

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 December 2001 (06.12.2001)

PCT

(10) International Publication Number
WO 01/92525 A2

(51) International Patent Classification⁷: C12N 15/12,
15/62, C07K 14/47, 16/30, C12Q 1/68, G01N 33/53,
33/574, A61K 39/00, 39/395, A61P 35/00

(21) International Application Number: PCT/US01/17066

(22) International Filing Date: 25 May 2001 (25.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/207,485 26 May 2000 (26.05.2000) US
60/230,475 6 September 2000 (06.09.2000) US

(71) Applicant (for all designated States except US): CORIXA
CORPORATION [US/US]; 1124 Columbia Street, Suite
200, Seattle, WA 98104 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): HARLOCKER,
Susan, L [US/US]; 7522 13th Avenue W, Seattle, WA
98117 (US). WANG, Tongtong [US/US]; 8049 N.E. 28th
Street, Medina, WA 98039 (US). BANGUR, Chaitanya,
S. [IN/US]; 2102 N. 105th Street, Apt. J101, Seattle, WA
98133 (US). KLEE, Jennifer, Irmele [US/US]; 5255 19th
Avenue N.E., Seattle, WA 98105 (US). SWITZER, Ann
[US/US]; 3838 N.E. 85th Street, Seattle, WA 98115 (US).

(74) Agents: Verna, James, V.; Seed Intellectual Property
Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle,
WA 98104-7092 et al. (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SI, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/92525 A2

(54) Title: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF LUNG CANCER

(57) Abstract: Compositions and methods for the therapy and diagnosis of cancer, particularly lung cancer, are disclosed. Illustrative compositions comprise one or more lung tumor polypeptides, immunogenic portions thereof, polynucleotides that encode such polypeptides, antigen presenting cell that expresses such polypeptides, and T cells that are specific for cells expressing such polypeptides. The disclosed compositions are useful, for example, in the diagnosis, prevention and/or treatment of diseases, particularly lung cancer.

COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS
OF LUNG CANCER

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to therapy and diagnosis of cancer, such as lung cancer. The invention is more specifically related to polypeptides, comprising at least a portion of a lung tumor protein, and to polynucleotides encoding such polypeptides. Such polypeptides and polynucleotides are useful in pharmaceutical compositions, e.g., vaccines, and other compositions for the diagnosis and treatment of lung cancer.

10 BACKGROUND OF THE INVENTION

Cancer is a significant health problem throughout the world. Although advances have been made in detection and therapy of cancer, no vaccine or other universally successful method for prevention and/or treatment is currently available. Current therapies, which are generally based on a combination of chemotherapy or 15 surgery and radiation, continue to prove inadequate in many patients.

Lung cancer is a significant health problem throughout the world. In the U.S., lung cancer is the primary cause of cancer death among both men and women, with an estimated 172,000 new cases being reported in 1994. The five-year survival rate among all lung cancer patients, regardless of the stage of disease at diagnosis, is 20 only 13%. This contrasts with a five-year survival rate of 46% among cases detected while the disease is still localized. However, early detection of lung cancer is difficult since clinical symptoms are often not seen until the disease has reached an advanced stage, and only 16% of lung cancers are discovered before the disease has spread.

In spite of considerable research into therapies for these and other 25 cancers, lung cancer remains difficult to diagnose and treat effectively. Accordingly, there is a need in the art for improved methods for detecting and treating such cancers. The present invention fulfills these needs and further provides other related advantages.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides polynucleotide compositions comprising a sequence selected from the group consisting of:

- (a) sequences provided in SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95;
- 5 (b) complements of the sequences provided in SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95;
- (c) sequences consisting of at least 20, 25, 30, 35, 40, 45, 50, 75 and 100 contiguous residues of a sequence provided in SEQ ID NO: 1-35, 42-55, 58-60, 63-91 and 93-95;
- 10 (d) sequences that hybridize to a sequence provided in SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95, under moderate or highly stringent conditions;
- (e) sequences having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity to a sequence of SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 15 93-95; and
- (f) degenerate variants of a sequence provided in SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95.

In one preferred embodiment, the polynucleotide compositions of the invention are expressed in at least about 20%, more preferably in at least about 30%, 20 and most preferably in at least about 50% of lung tumors samples tested, at a level that is at least about 2-fold, preferably at least about 5-fold, and most preferably at least about 10-fold higher than that for normal tissues.

The present invention, in another aspect, provides polypeptide compositions comprising an amino acid sequence that is encoded by a polynucleotide 25 sequence described above.

The present invention further provides polypeptide compositions comprising an amino acid sequence selected from the group consisting of sequences recited in SEQ ID NO:36-41, 56, 57, 61, 62, 92 and 96.

In certain preferred embodiments, the polypeptides and/or 30 polynucleotides of the present invention are immunogenic, *i.e.*, they are capable of

eliciting an immune response, particularly a humoral and/or cellular immune response, as further described herein.

The present invention further provides fragments, variants and/or derivatives of the disclosed polypeptide and/or polynucleotide sequences, wherein the
5 fragments, variants and/or derivatives preferably have a level of immunogenic activity of at least about 50%, preferably at least about 70% and more preferably at least about 90% of the level of immunogenic activity of a polypeptide sequence set forth in SEQ ID NO:36-41, 56, 57, 61, 62, 92 and 96 or a polypeptide sequence encoded by a polynucleotide sequence set forth in SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95.

10 The present invention further provides polynucleotides that encode a polypeptide described above, expression vectors comprising such polynucleotides and host cells transformed or transfected with such expression vectors.

Within other aspects, the present invention provides pharmaceutical compositions comprising a polypeptide or polynucleotide as described above and a
15 physiologically acceptable carrier.

Within a related aspect of the present invention, the pharmaceutical compositions, *e.g.*, vaccine compositions, are provided for prophylactic or therapeutic applications. Such compositions generally comprise an immunogenic polypeptide or polynucleotide of the invention and an immunostimulant, such as an adjuvant.

20 The present invention further provides pharmaceutical compositions that comprise: (a) an antibody or antigen-binding fragment thereof that specifically binds to a polypeptide of the present invention, or a fragment thereof; and (b) a physiologically acceptable carrier.

Within further aspects, the present invention provides pharmaceutical compositions comprising: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) a pharmaceutically acceptable carrier or excipient. Illustrative antigen presenting cells include dendritic cells, macrophages, monocytes, fibroblasts and B cells.

Within related aspects, pharmaceutical compositions are provided that
30 comprise: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) an immunostimulant.

The present invention further provides, in other aspects, fusion proteins that comprise at least one polypeptide as described above, as well as polynucleotides encoding such fusion proteins, typically in the form of pharmaceutical compositions, e.g., vaccine compositions, comprising a physiologically acceptable carrier and/or an immunostimulant. The fusions proteins may comprise multiple immunogenic polypeptides or portions/variants thereof, as described herein, and may further comprise one or more polypeptide segments for facilitating the expression, purification and/or immunogenicity of the polypeptide(s).

Within further aspects, the present invention provides methods for stimulating an immune response in a patient, preferably a T cell response in a human patient, comprising administering a pharmaceutical composition described herein. The patient may be afflicted with lung cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient a pharmaceutical composition as recited above. The patient may be afflicted with lung cancer, in which case the methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

The present invention further provides, within other aspects, methods for removing tumor cells from a biological sample, comprising contacting a biological sample with T cells that specifically react with a polypeptide of the present invention, wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological sample treated as described above.

Methods are further provided, within other aspects, for stimulating and/or expanding T cells specific for a polypeptide of the present invention, comprising contacting T cells with one or more of: (i) a polypeptide as described above; (ii) a polynucleotide encoding such a polypeptide; and/or (iii) an antigen presenting cell that

expresses such a polypeptide; under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. Isolated T cell populations comprising T cells prepared as described above are also provided.

Within further aspects, the present invention provides methods for
5 inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population as described above.

The present invention further provides methods for inhibiting the development of a cancer in a patient, comprising the steps of: (a) incubating CD4⁺ and/or CD8⁺ T cells isolated from a patient with one or more of: (i) a polypeptide
10 comprising at least an immunogenic portion of polypeptide disclosed herein; (ii) a polynucleotide encoding such a polypeptide; and (iii) an antigen-presenting cell that expressed such a polypeptide; and (b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient. Proliferated cells may, but need not, be cloned prior to administration to the
15 patient.

Within further aspects, the present invention provides methods for determining the presence or absence of a cancer, preferably a lung cancer, in a patient comprising: (a) contacting a biological sample obtained from a patient with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount
20 of polypeptide that binds to the binding agent; and (c) comparing the amount of polypeptide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within preferred embodiments, the binding agent is an antibody, more preferably a monoclonal antibody.

The present invention also provides, within other aspects, methods for monitoring the progression of a cancer in a patient. Such methods comprise the steps of: (a) contacting a biological sample obtained from a patient at a first point in time with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in
25 time; and (d) comparing the amount of polypeptide detected in step (c) with the amount
30

detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

- The present invention further provides, within other aspects, methods for determining the presence or absence of a cancer in a patient, comprising the steps of: (a)
- 5 contacting a biological sample, e.g., tumor sample, serum sample, etc., obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample a level of a polynucleotide, preferably mRNA, that hybridizes to the oligonucleotide; and (c) comparing the level of polynucleotide that hybridizes to the oligonucleotide with a
 - 10 predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction using, for example, at least one oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such a polynucleotide. Within other embodiments, the amount of
 - 15 mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

- In related aspects, methods are provided for monitoring the progression of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

- Within further aspects, the present invention provides antibodies, such as monoclonal antibodies, that bind to a polypeptide as described above, as well as diagnostic kits comprising such antibodies. Diagnostic kits comprising one or more
- 30 oligonucleotide probes or primers as described above are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

5 SEQUENCE IDENTIFIERS

SEQ ID NO:1 is the cDNA sequence for Clone ID # 55964 which is named clone L1040C, and is the same sequence as SEQ ID NO:2337 from U.S. Provisional Application 60/207,485.

SEQ ID NO:2 is an extended cDNA sequence for L1040C (Clone ID # 10 55964).

SEQ ID NO:3 is the cDNA sequence for Clone ID # 58269 which is named clone L1039C, and is the same sequence as SEQ ID NO:7264 from U.S. Provisional Application 60/207,485.

SEQ ID NO:4 is an extended cDNA sequence for L1039C (Clone ID # 15 58269), and which corresponds to the fbx5 F-box gene.

SEQ ID NO:5 is the cDNA sequence for Clone ID # 58267 which is named clone L1037C, and is the same sequence as SEQ ID NO:4978 from U.S. Provisional Application 60/207,485.

SEQ ID NO:6 is an extended cDNA sequence for L1037C (Clone ID # 20 58267), and which corresponds to the mitotic checkpoint kinase mad3-like gene.

SEQ ID NO:7 is the cDNA sequence for Clone ID # 58245 which is named clone L1038C, and is the same sequence as SEQ ID NO:1796 from U.S. Provisional Application 60/207,485.

SEQ ID NO:8 is an extended cDNA sequence for L1038C (Clone ID # 25 58245), and which corresponds to a neuronal ER localized gene.

SEQ ID NO:9 is the cDNA sequence for Clone ID # 55571 which is named clone L1027C, and is the same sequence as SEQ ID NO:4538 from U.S. Provisional Application 60/207,485.

SEQ ID NO:10 is an extended cDNA sequence for L1027C (Clone ID # 30 55571).

SEQ ID NO:11 is the cDNA sequence for Clone ID # 55978.

SEQ ID NO:12 is an extended cDNA sequence for Clone ID # 55978.

SEQ ID NO:13 is the cDNA sequence for Clone ID # 55980.

SEQ ID NO:14 is an extended cDNA sequence for Clone ID # 55980.

5 SEQ ID NO:15 is the cDNA sequence for Clone ID # 58346.

SEQ ID NO:16 is an extended cDNA sequence for Clone ID # 58346.

SEQ ID NO:17 is the cDNA sequence for Clone ID # 55561.

SEQ ID NO:18 is an extended cDNA sequence for Clone ID # 55561.

SEQ ID NO:19 is the cDNA sequence for Clone ID # 55984.

10 SEQ ID NO:20 is an extended cDNA sequence for Clone ID # 55984,
and which corresponds to a gt mismatch glycosylase gene.

SEQ ID NO:21 is the cDNA sequence for Clone ID # 58261.

SEQ ID NO:22 is an extended cDNA sequence for Clone ID # 58261,
and which corresponds to a phosphoserine aminotransferase gene.

15 SEQ ID NO:23 is the cDNA sequence for Clone ID # 58348.

SEQ ID NO:24 is an extended cDNA sequence for Clone ID # 58348,
and which corresponds to a hCAP gene.

SEQ ID NO:25 is the cDNA sequence for Clone ID # 56016.

SEQ ID NO:26 is an extended cDNA sequence for Clone ID # 56016.

20 SEQ ID NO:27 is the cDNA sequence for Clone ID # 55987.

SEQ ID NO:28 is an extended cDNA sequence for Clone ID # 55987.

SEQ ID NO:29 is the cDNA sequence for Clone ID # 55956.

SEQ ID NO:30 is an extended cDNA sequence for Clone ID # 55956.

SEQ ID NO:31 is the cDNA sequence for Clone ID # 55952.

25 SEQ ID NO:32 is the cDNA sequence for Clone ID # 55957.

SEQ ID NO:33 is an extended cDNA sequence for Clone ID # 55957.

SEQ ID NO:34 is the cDNA sequence for Clone ID # 55559.

SEQ ID NO:35 is an extended cDNA sequence for Clone ID # 55559.

30 SEQ ID NO:36 is an amino acid sequence of an ORF for L1027C,
encoded by the polynucleotide of SEQ ID NO: 10.

SEQ ID NO:37 is an amino acid sequence of the F-box protein Fbx5 encoded by SEQ ID NO:4.

SEQ ID NO:38 is an amino acid sequence of the mitotic checkpoint kinase MAD3-like protein encoded by SEQ ID NO:6.

5 SEQ ID NO:39 is an amino acid sequence of the neuronal olfactomedin-related ER localized protein encoded by SEQ ID NO:8.

SEQ ID NO:40 is an amino acid sequence of the phosphoserine aminotransferase encoded by SEQ ID NO:22.

10 SEQ ID NO:41 is an amino acid sequence of the gt mismatch glycosylase encoded by SEQ ID NO:20.

SEQ ID NO:42 is the determined cDNA sequence for Clone ID # 63575 which is named clone L1053C.

SEQ ID NO:43 is the determined cDNA sequence for Clone ID # 63582 which is named clone L1054C.

15 SEQ ID NO:44 is the determined cDNA sequence for Clone ID # 63598 which is named clone L1055C.

SEQ ID NO:45 is the determined cDNA sequence for Clone ID # 64963 which is named clone L1056C.

20 SEQ ID NO:46 is the determined cDNA sequence for Clone ID # 64988 which is named clone L1058C.

SEQ ID NO:47 is the determined cDNA sequence for Clone ID # 63485.

SEQ ID NO:48 is the determined cDNA sequence for Clone ID # 65010.

25 SEQ ID NO:49 is a predicted full-length cDNA sequence for SEQ ID NO:42 which is a full-length sequence from Genbank for an insulinoma-associated 1 mRNA.

SEQ ID NO:50 is a predicted full-length cDNA sequence for SEQ ID NO:43 which is a full-length sequence from Genbank for KIAA0535.

SEQ ID NO:51 is a predicted extended cDNA sequence for SEQ ID NO:44.

30 SEQ ID NO:52 is a predicted full-length cDNA sequence for SEQ ID NO:45 which is a full-length sequence from genbank for a human DAZ mRNA 3'UTR.

SEQ ID NO:53 is a predicted extended cDNA sequence for SEQ ID NO:46.

SEQ ID NO:54 is a predicted extended cDNA sequence for SEQ ID NO:47.

5 SEQ ID NO:55 is a predicted extended cDNA sequence for SEQ ID NO:48.

SEQ ID NO:56 is the deduced amino acid sequence encoded by SEQ ID NO:49.

10 SEQ ID NO:57 is the deduced amino acid sequence encoded by SEQ ID NO:50.

SEQ ID NO:58 is the determined full-length cDNA sequence for clone L1058C (sequence of the originally isolated clone is given in SEQ ID NO:46 and the predicted extended cDNA sequence in SEQ ID NO:53).

SEQ ID NO:59 is a first predicted ORF of SEQ ID NO:58.

15 SEQ ID NO:60 is a second predicted ORF of SEQ ID NO:58.

SEQ ID NO:61 is the deduced amino acid sequence encoded by SEQ ID NO:59.

SEQ ID NO:62 is the deduced amino acid sequence encoded by SEQ ID NO:60.

20 SEQ ID NO:63 is the determined cDNA sequence for Clone ID # 72761.

SEQ ID NO:64 is the determined cDNA sequence for Clone ID # 72762.

SEQ ID NO:65 is the determined cDNA sequence for Clone ID # 72763.

SEQ ID NO:66 is the determined cDNA sequence for Clone ID # 72764.

SEQ ID NO:67 is the determined cDNA sequence for Clone ID # 72765.

25 SEQ ID NO:68 is the determined cDNA sequence for Clone ID # 72766.

SEQ ID NO:69 is the determined cDNA sequence for Clone ID # 72772.

SEQ ID NO:70 is the determined cDNA sequence for Clone ID # 72775.

SEQ ID NO:71 is the determined cDNA sequence for Clone ID # 72776.

SEQ ID NO:72 is the determined cDNA sequence for Clone ID # 72779.

30 SEQ ID NO:73 is the determined cDNA sequence for Clone ID # 72781.

SEQ ID NO:74 is the determined cDNA sequence for Clone ID # 72784.

SEQ ID NO:75 is the determined cDNA sequence for Clone ID # 72788.
SEQ ID NO:76 is the determined cDNA sequence for Clone ID # 72789.
SEQ ID NO:77 is the determined cDNA sequence for Clone ID # 72790.
SEQ ID NO:78 is the determined cDNA sequence for Clone ID # 72791.
5 SEQ ID NO:79 is the determined cDNA sequence for Clone ID # 72792.
SEQ ID NO:80 is the determined cDNA sequence for Clone ID # 72794.
SEQ ID NO:81 is the determined cDNA sequence for Clone ID # 72795.
SEQ ID NO:82 is the determined cDNA sequence for Clone ID # 72797.
SEQ ID NO:83 is the determined cDNA sequence for Clone ID # 72798.
10 SEQ ID NO:84 is the determined cDNA sequence for Clone ID # 72804.
SEQ ID NO:85 is the determined cDNA sequence for Clone ID # 72805.
SEQ ID NO:86 is the determined cDNA sequence for Clone ID # 72806.
SEQ ID NO:87 is the determined cDNA sequence for Clone ID # 72807.
SEQ ID NO:88 is the determined cDNA sequence for Clone ID # 72808.
15 SEQ ID NO:89 is the determined cDNA sequence for Clone ID # 72809.
SEQ ID NO:90 is the determined cDNA sequence for Clone ID # 72811.
SEQ ID NO:91 is the determined full-length cDNA sequence for Clone
ID # 72813 which is named clone L1080C.
SEQ ID NO:92 is the deduced amino acid sequence encoded by SEQ ID
20 NO:91.
SEQ ID NO:93 is the ORF for L1027C.
SEQ ID NO:94 is a first determined full-length cDNA sequence for
L1027C.
SEQ ID NO:95 is a second determined full-length cDNA sequence for
25 L1027C.
SEQ ID NO:96 is the deduced amino acid sequence encoded by SEQ ID
NO:93.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed generally to compositions and their use
30 in the therapy and diagnosis of cancer, particularly lung cancer. As described further

below, illustrative compositions of the present invention include, but are not restricted to, polypeptides, particularly immunogenic polypeptides, polynucleotides encoding such polypeptides, antibodies and other binding agents, antigen presenting cells (APCs) and immune system cells (*e.g.*, T cells).

- 5 The practice of the present invention will employ, unless indicated specifically to the contrary, conventional methods of virology, immunology, microbiology, molecular biology and recombinant DNA techniques within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook, et al. Molecular
10 Cloning: A Laboratory Manual (2nd Edition, 1989); Maniatis et al. Molecular Cloning: A Laboratory Manual (1982); DNA Cloning: A Practical Approach, vol. I & II (D. Glover, ed.); Oligonucleotide Synthesis (N. Gait, ed., 1984); Nucleic Acid Hybridization (B. Hames & S. Higgins, eds., 1985); Transcription and Translation (B. Hames & S. Higgins, eds., 1984); Animal Cell Culture (R. Freshney, ed., 1986); Perbal,
15 A Practical Guide to Molecular Cloning (1984).

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

- As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates
20 otherwise.

Polypeptide Compositions

- As used herein, the term "polypeptide" is used in its conventional meaning, *i.e.*, as a sequence of amino acids. The polypeptides are not limited to a specific length of the product; thus, peptides, oligopeptides, and proteins are included
25 within the definition of polypeptide, and such terms may be used interchangeably herein unless specifically indicated otherwise. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like, as well as other modifications known in the art, both naturally occurring and non-naturally occurring. A polypeptide may be an entire
30 protein, or a subsequence thereof. Particular polypeptides of interest in the context of

this invention are amino acid subsequences comprising epitopes, *i.e.*, antigenic determinants substantially responsible for the immunogenic properties of a polypeptide and being capable of evoking an immune response.

- Particularly illustrative polypeptides of the present invention comprise
- 5 those encoded by a polynucleotide sequence set forth in any one of SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95, or a sequence that hybridizes under moderately stringent conditions, or, alternatively, under highly stringent conditions, to a polynucleotide sequence set forth in any one of SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95. Certain other illustrative polypeptides of the invention comprise amino acid sequences
- 10 as set forth in any one of SEQ ID NOs:36-41, 56, 57, 61, 62, 92 and 96.

The polypeptides of the present invention are sometimes herein referred to as lung tumor proteins or lung tumor polypeptides, as an indication that their identification has been based at least in part upon their increased levels of expression in lung tumor samples. Thus, a "lung tumor polypeptide" or "lung tumor protein," refers

15 generally to a polypeptide sequence of the present invention, or a polynucleotide sequence encoding such a polypeptide, that is expressed in a substantial proportion of lung tumor samples, for example preferably greater than about 20%, more preferably greater than about 30%, and most preferably greater than about 50% or more of lung tumor samples tested, at a level that is at least two fold, and preferably at least five fold,

20 greater than the level of expression in normal tissues, as determined using a representative assay provided herein. A lung tumor polypeptide sequence of the invention, based upon its increased level of expression in tumor cells, has particular utility both as a diagnostic marker as well as a therapeutic target, as further described below.

25 In certain preferred embodiments, the polypeptides of the invention are immunogenic, *i.e.*, they react detectably within an immunoassay (such as an ELISA or T-cell stimulation assay) with antisera and/or T-cells from a patient with lung cancer. Screening for immunogenic activity can be performed using techniques well known to the skilled artisan. For example, such screens can be performed using methods such as

30 those described in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one illustrative example, a polypeptide may be

immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, ¹²⁵I-labeled Protein A.

As would be recognized by the skilled artisan, immunogenic portions of the polypeptides disclosed herein are also encompassed by the present invention. An "immunogenic portion," as used herein, is a fragment of an immunogenic polypeptide of the invention that itself is immunologically reactive (*i.e.*, specifically binds) with the B-cells and/or T-cell surface antigen receptors that recognize the polypeptide. Immunogenic portions may generally be identified using well known techniques, such 10 as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (*i.e.*, they react with the protein in an ELISA or other 15 immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well-known techniques.

In one preferred embodiment, an immunogenic portion of a polypeptide of the present invention is a portion that reacts with antisera and/or T-cells at a level that is not substantially less than the reactivity of the full-length polypeptide (*e.g.*, in an 20 ELISA and/or T-cell reactivity assay). Preferably, the level of immunogenic activity of the immunogenic portion is at least about 50%, preferably at least about 70% and most preferably greater than about 90% of the immunogenicity for the full-length polypeptide. In some instances, preferred immunogenic portions will be identified that have a level of immunogenic activity greater than that of the corresponding full-length 25 polypeptide, *e.g.*, having greater than about 100% or 150% or more immunogenic activity.

In certain other embodiments, illustrative immunogenic portions may include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other illustrative immunogenic portions will contain a small N- 30 and/or C-terminal deletion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

In another embodiment, a polypeptide composition of the invention may also comprise one or more polypeptides that are immunologically reactive with T cells and/or antibodies generated against a polypeptide of the invention, particularly a polypeptide having an amino acid sequence disclosed herein, or to an immunogenic fragment or variant thereof.

In another embodiment of the invention, polypeptides are provided that comprise one or more polypeptides that are capable of eliciting T cells and/or antibodies that are immunologically reactive with one or more polypeptides described herein, or one or more polypeptides encoded by contiguous nucleic acid sequences contained in 10 the polynucleotide sequences disclosed herein, or immunogenic fragments or variants thereof, or to one or more nucleic acid sequences which hybridize to one or more of these sequences under conditions of moderate to high stringency.

The present invention, in another aspect, provides polypeptide fragments comprising at least about 5, 10, 15, 20, 25, 50, or 100 contiguous amino acids, or more, 15 including all intermediate lengths, of a polypeptide compositions set forth herein, such as those set forth in SEQ ID NOS:36-41, 56, 57, 61, 62, 92 and 96, or those encoded by a polynucleotide sequence set forth in a sequence of SEQ ID NOS:1-35, 42-55, 58-60, 63-91 and 93-95.

In another aspect, the present invention provides variants of the 20 polypeptide compositions described herein. Polypeptide variants generally encompassed by the present invention will typically exhibit at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity (determined as described below), along its length, to a polypeptide sequences set forth herein.

25 In one preferred embodiment, the polypeptide fragments and variants provided by the present invention are immunologically reactive with an antibody and/or T-cell that react with a full-length polypeptide specifically set forth herein.

In another preferred embodiment, the polypeptide fragments and variants provided by the present invention exhibit a level of immunogenic activity of at least 30 about 50%, preferably at least about 70%, and most preferably at least about 90% or

more of that exhibited by a full-length polypeptide sequence specifically set forth herein.

A polypeptide "variant," as the term is used herein, is a polypeptide that typically differs from a polypeptide specifically disclosed herein in one or more substitutions, deletions, additions and/or insertions. Such variants may be naturally occurring or may be synthetically generated, for example, by modifying one or more of the above polypeptide sequences of the invention and evaluating their immunogenic activity as described herein and/or using any of a number of techniques well known in the art.

10 For example, certain illustrative variants of the polypeptides of the invention include those in which one or more portions, such as an N-terminal leader sequence or transmembrane domain, have been removed. Other illustrative variants include variants in which a small portion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids) has been removed from the N- and/or C-terminal of the mature protein.

15 In many instances, a variant will contain conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydrophobic nature of the polypeptide to be substantially unchanged. As described above, modifications may be
20 made in the structure of the polynucleotides and polypeptides of the present invention and still obtain a functional molecule that encodes a variant or derivative polypeptide with desirable characteristics, *e.g.*, with immunogenic characteristics. When it is desired to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, immunogenic variant or portion of a polypeptide of the invention,
25 one skilled in the art will typically change one or more of the codons of the encoding DNA sequence according to Table 1.

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites
30 on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence

substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides

5 without appreciable loss of their biological utility or activity.

Table 1

| Amino Acids | | Codons | | | | | |
|---------------|-----|--------|-----|-----|-----|-----|-----|
| Alanine | Ala | A | GCA | GCC | GCG | GCU | |
| Cysteine | Cys | C | UGC | UGU | | | |
| Aspartic acid | Asp | D | GAC | GAU | | | |
| Glutamic acid | Glu | E | GAA | GAG | | | |
| Phenylalanine | Phe | F | UUC | UUU | | | |
| Glycine | Gly | G | GGA | GGC | GGG | GGU | |
| Histidine | His | H | CAC | CAU | | | |
| Isoleucine | Ile | I | AUA | AUC | AUU | | |
| Lysine | Lys | K | AAA | AAG | | | |
| Leucine | Leu | L | UUA | UUG | CUA | CUC | CUG |
| Methionine | Met | M | AUG | | | | |
| Asparagine | Asn | N | AAC | AAU | | | |
| Proline | Pro | P | CCA | CCC | CCG | CCU | |
| Glutamine | Gln | Q | CAA | CAG | | | |
| Arginine | Arg | R | AGA | AGG | CGA | CGC | CGG |
| Serine | Ser | S | AGC | AGU | UCA | UCC | UCG |
| Threonine | Thr | T | ACA | ACC | ACG | ACU | |
| Valine | Val | V | GUU | GUC | GUG | GUU | |
| Tryptophan | Trp | W | UGG | | | | |
| Tyrosine | Tyr | Y | UAC | UAU | | | |

In making such changes, the hydrophobic index of amino acids may be
10 considered. The importance of the hydrophobic amino acid index in conferring

interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporated herein by reference). It is accepted that the relative hydrophobic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been assigned a hydrophobic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982). These values are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydrophobic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydrophobic indices are within ± 2 is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U. S. Patent 4,554,101 (specifically incorporated herein by reference in its entirety), states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate ($+3.0 \pm 1$); glutamate ($+3.0 \pm 1$); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 ± 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ± 2

is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

In addition, any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of nontraditional bases such as inosine, queosine and wybutoxine, as well as acetyl-methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

Amino acid substitutions may further be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may represent conservative changes include: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or alternatively, contain nonconservative changes. In a preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity, secondary structure and hydrophobic nature of the polypeptide.

As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide 5 (e.g., poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

When comparing polypeptide sequences, two sequences are said to be "identical" if the sequence of amino acids in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two 10 sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences 15 are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A 20 model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) 25 *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Saitou, N. Nei, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad. Sci. USA* 80:726-730.

30 Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL.*

- Math 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. For amino acid sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment.

In one preferred approach, the “percentage of sequence identity” is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Within other illustrative embodiments, a polypeptide may be a xenogeneic polypeptide that comprises an polypeptide having substantial sequence identity, as described above, to the human polypeptide (also termed autologous antigen) which served as a reference polypeptide, but which xenogeneic polypeptide is derived 5 from a different, non-human species. One skilled in the art will recognize that "self" antigens are often poor stimulators of CD8+ and CD4+ T-lymphocyte responses, and therefore efficient immunotherapeutic strategies directed against tumor polypeptides require the development of methods to overcome immune tolerance to particular self tumor polypeptides. For example, humans immunized with prostase 10 protein from a xenogeneic (non human) origin are capable of mounting an immune response against the counterpart human protein, e.g. the human prostase tumor protein present on human tumor cells. Accordingly, the present invention provides methods for purifying the xenogeneic form of the tumor proteins set forth herein, such as the polypeptides set forth in SEQ ID NO:36-41, 56, 57, 61, 62, 92 and 96, or those encoded 15 by polynucleotide sequences set forth in SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95.

Therefore, one aspect of the present invention provides xenogeneic variants of the polypeptide compositions described herein. Such xenogeneic variants generally encompassed by the present invention will typically exhibit at least about 20 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity along their lengths, to a polypeptide sequences set forth herein.

More particularly, the invention is directed to mouse, rat, monkey, porcine and other non-human polypeptides which can be used as xenogeneic forms of human polypeptides set forth herein, to induce immune responses directed against 25 tumor polypeptides of the invention.

Within other illustrative embodiments, a polypeptide may be a fusion polypeptide that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes 30 (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the

native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the polypeptide or to enable the polypeptide to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, 5 which facilitate purification of the polypeptide.

Fusion polypeptides may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion polypeptide is expressed as a recombinant polypeptide, allowing the production of increased levels, relative to a non-fused polypeptide, in an expression system. Briefly, DNA sequences 10 encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion polypeptide 15 that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion polypeptide using standard techniques well known in the 20 art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, 25 Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino 30 acids in length. Linker sequences are not required when the first and second

polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements 5 responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

The fusion polypeptide can comprise a polypeptide as described herein 10 together with an unrelated immunogenic protein, such as an immunogenic protein capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (see, for example, Stoute *et al.* *New Engl. J. Med.*, 336:86-91, 1997).

In one preferred embodiment, the immunological fusion partner is 15 derived from a *Mycobacterium* sp., such as a *Mycobacterium tuberculosis*-derived Ra12 fragment. Ra12 compositions and methods for their use in enhancing the expression and/or immunogenicity of heterologous polynucleotide/polypeptide sequences is described in U.S. Patent Application 60/158,585, the disclosure of which is incorporated herein by reference in its entirety. Briefly, Ra12 refers to a polynucleotide 20 region that is a subsequence of a *Mycobacterium tuberculosis* MTB32A nucleic acid. MTB32A is a serine protease of 32 KD molecular weight encoded by a gene in virulent and avirulent strains of *M. tuberculosis*. The nucleotide sequence and amino acid sequence of MTB32A have been described (for example, U.S. Patent Application 60/158,585; see also, Skeiky *et al.*, *Infection and Immun.* (1999) 67:3998-4007, 25 incorporated herein by reference). C-terminal fragments of the MTB32A coding sequence express at high levels and remain as a soluble polypeptides throughout the purification process. Moreover, Ra12 may enhance the immunogenicity of heterologous immunogenic polypeptides with which it is fused. One preferred Ra12 fusion . polypeptide comprises a 14 KD C-terminal fragment corresponding to amino acid 30 residues 192 to 323 of MTB32A. Other preferred Ra12 polynucleotides generally comprise at least about 15 consecutive nucleotides, at least about 30 nucleotides, at least

about 60 nucleotides, at least about 100 nucleotides, at least about 200 nucleotides, or at least about 300 nucleotides that encode a portion of a Ra12 polypeptide. Ra12 polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a Ra12 polypeptide or a portion thereof) or may comprise a variant of such a sequence. Ra12 polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded fusion polypeptide is not substantially diminished, relative to a fusion polypeptide comprising a native Ra12 polypeptide. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90% identity to a polynucleotide sequence that encodes a native Ra12 polypeptide or a portion thereof.

Within other preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenzae* B (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemagglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the LytA gene; *Gene* 43:265-292, 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for

expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (*see Biotechnology* 10:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion polypeptide. A repeat portion is found in the C-terminal region starting at 5 residue 178. A particularly preferred repeat portion incorporates residues 188-305.

Yet another illustrative embodiment involves fusion polypeptides, and the polynucleotides encoding them, wherein the fusion partner comprises a targeting signal capable of directing a polypeptide to the endosomal/lysosomal compartment, as described in U.S. Patent No. 5,633,234. An immunogenic polypeptide of the invention, 10 when fused with this targeting signal, will associate more efficiently with MHC class II molecules and thereby provide enhanced *in vivo* stimulation of CD4⁺ T-cells specific for the polypeptide.

Polypeptides of the invention are prepared using any of a variety of well known synthetic and/or recombinant techniques, the latter of which are further 15 described below. Polypeptides, portions and other variants generally less than about 150 amino acids can be generated by synthetic means, using techniques well known to those of ordinary skill in the art. In one illustrative example, such polypeptides are synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a 20 growing amino acid chain. *See Merrifield, J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

In general, polypeptide compositions (including fusion polypeptides) of 25 the invention are isolated. An "isolated" polypeptide is one that is removed from its original environment. For example, a naturally-occurring protein or polypeptide is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are also purified, *e.g.*, are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% 30 pure.

Polynucleotide Compositions

The present invention, in other aspects, provides polynucleotide compositions. The terms "DNA" and "polynucleotide" are used essentially interchangeably herein to refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. "Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA molecule does not contain large portions of unrelated coding DNA, such as large chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA molecule as originally isolated, and does not exclude 10 genes or coding regions later added to the segment by the hand of man.

As will be understood by those skilled in the art, the polynucleotide compositions of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, peptides and the like. Such segments may 15 be naturally isolated, or modified synthetically by the hand of man.

As will be also recognized by the skilled artisan, polynucleotides of the invention may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules may include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-20 to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous 25 sequence that encodes a polypeptide/protein of the invention or a portion thereof) or may comprise a sequence that encodes a variant or derivative, preferably and immunogenic variant or derivative, of such a sequence.

Therefore, according to another aspect of the present invention, polynucleotide compositions are provided that comprise some or all of a polynucleotide 30 sequence set forth in any one of SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95, complements of a polynucleotide sequence set forth in any one of SEQ ID NO:1-35, 42-

55, 58-60, 63-91 and 93-95, and degenerate variants of a polynucleotide sequence set forth in any one of SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95. In certain preferred embodiments, the polynucleotide sequences set forth herein encode immunogenic polypeptides, as described above.

5 In other related embodiments, the present invention provides polynucleotide variants having substantial identity to the sequences disclosed herein in SEQ ID NO:1-35, 42-55, 58-60, 63-91 and 93-95, for example those comprising at least 70% sequence identity, preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or higher, sequence identity compared to a polynucleotide sequence of this 10 invention using the methods described herein, (e.g., BLAST analysis using standard parameters, as described below). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

15 Typically, polynucleotide variants will contain one or more substitutions, additions, deletions and/or insertions, preferably such that the immunogenicity of the polypeptide encoded by the variant polynucleotide is not substantially diminished relative to a polypeptide encoded by a polynucleotide sequence specifically set forth herein). The term "variants" should also be understood to encompasses homologous 20 genes of xenogenic origin.

In additional embodiments, the present invention provides polynucleotide fragments comprising or consisting of various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that 25 comprise or consist of at least about 10, 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It will be readily understood that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, *etc.*; 21, 22, 23, *etc.*; 30, 31, 32, *etc.*; 50, 51, 52, 53, *etc.*; 30 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through 200-500; 500-1,000, and the like. A polynucleotide sequence as described here may be

extended at one or both ends by additional nucleotides not found in the native sequence. This additional sequence may consist of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides at either end of the disclosed sequence or at both ends of the disclosed sequence.

- 5 In another embodiment of the invention, polynucleotide compositions are provided that are capable of hybridizing under moderate to high stringency conditions to a polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for
- 10 testing the hybridization of a polynucleotide of this invention with other polynucleotides include prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-60°C, 5 X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS. One skilled in the art will understand that the stringency of hybridization can be readily manipulated,
- 15 such as by altering the salt content of the hybridization solution and/or the temperature at which the hybridization is performed. For example, in another embodiment, suitable highly stringent hybridization conditions include those described above, with the exception that the temperature of hybridization is increased, *e.g.*, to 60-65°C or 65-70°C.
- 20 In certain preferred embodiments, the polynucleotides described above, *e.g.*, polynucleotide variants, fragments and hybridizing sequences, encode polypeptides that are immunologically cross-reactive with a polypeptide sequence specifically set forth herein. In other preferred embodiments, such polynucleotides encode polypeptides that have a level of immunogenic activity of at least about 50%, preferably
- 25 at least about 70%, and more preferably at least about 90% of that for a polypeptide sequence specifically set forth herein.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment

- of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative polynucleotide segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 5 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

When comparing polynucleotide sequences, two sequences are said to be "identical" if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two 10 sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences 15 are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A 20 model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) 25 *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad. Sci. USA* 80:726-730.

30 Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL.*

- Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics 5 Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 10 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative example, cumulative scores can be calculated using, for 15 nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; 20 or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and 25 a comparison of both strands.

Preferably, the “percentage of sequence identity” is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 30 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The

percentage is calculated by determining the number of positions at which the identical nucleic acid bases occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the
5 percentage of sequence identity.

It will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides
10 that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not,
15 have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

Therefore, in another embodiment of the invention, a mutagenesis approach, such as site-specific mutagenesis, is employed for the preparation of immunogenic variants and/or derivatives of the polypeptides described herein. By this
20 approach, specific modifications in a polypeptide sequence can be made through mutagenesis of the underlying polynucleotides that encode them. These techniques provides a straightforward approach to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the polynucleotide.

Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Mutations may be employed in a
30 selected polynucleotide sequence to improve, alter, decrease, modify, or otherwise

change the properties of the polynucleotide itself, and/or alter the properties, activity, composition, stability, or primary sequence of the encoded polypeptide.

- In certain embodiments of the present invention, the inventors contemplate the mutagenesis of the disclosed polynucleotide sequences to alter one or
- 5 more properties of the encoded polypeptide, such as the immunogenicity of a polypeptide vaccine. The techniques of site-specific mutagenesis are well-known in the art, and are widely used to create variants of both polypeptides and polynucleotides. For example, site-specific mutagenesis is often used to alter a specific portion of a DNA molecule. In such embodiments, a primer comprising typically about 14 to about 25
- 10 nucleotides or so in length is employed, with about 5 to about 10 residues on both sides of the junction of the sequence being altered.

As will be appreciated by those of skill in the art, site-specific mutagenesis techniques have often employed a phage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis

15 include vectors such as the M13 phage. These phage are readily commercially-available and their use is generally well-known to those skilled in the art. Double-stranded plasmids are also routinely employed in site directed mutagenesis that eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is

20 performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector that includes within its sequence a DNA sequence that encodes the desired peptide. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I

25 Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

30 The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis provides a means of producing

potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants. Specific details regarding these methods and protocols are found in the teachings of Maloy *et al.*, 1994; Segal, 1976; Prokop and Bajpai, 1991; Kuby, 1994; and Maniatis *et al.*, 1982, each incorporated herein by reference, for that purpose.

As used herein, the term "oligonucleotide directed mutagenesis procedure" refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template dependent process refers to nucleic acid synthesis of a RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U. S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety.

In another approach for the production of polypeptide variants of the present invention, recursive sequence recombination, as described in U.S. Patent No. 5,837,458, may be employed. In this approach, iterative cycles of recombination and screening or selection are performed to "evolve" individual polynucleotide variants of the invention having, for example, enhanced immunogenic activity.

In other embodiments of the present invention, the polynucleotide sequences provided herein can be advantageously used as probes or primers for nucleic acid hybridization. As such, it is contemplated that nucleic acid segments that comprise or consist of a sequence region of at least about a 15 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 15 nucleotide long

contiguous sequence disclosed herein will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 500, 1000 (including all intermediate lengths) and even up to full length sequences will also be of use in certain embodiments.

5 The ability of such nucleic acid probes to specifically hybridize to a sequence of interest will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are also envisioned, such as the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

10 Polynucleotide molecules having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide sequence disclosed herein, are particularly contemplated as hybridization probes for use in, *e.g.*, Southern and Northern blotting. This would allow a gene
15 product, or fragment thereof, to be analyzed, both in diverse cell types and also in various bacterial cells. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region
20 may be varied, such as between about 15 and about 100 nucleotides, but larger contiguous complementarity stretches may be used, according to the length complementary sequences one wishes to detect.

 The use of a hybridization probe of about 15-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules
25 having contiguous complementary sequences over stretches greater than 15 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where
30 desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequences set forth herein, or to any continuous portion of the sequences, from about 15-25 nucleotides in length up to and including the full length sequence, that one wishes to utilize as a probe 5 or primer. The choice of probe and primer sequences may be governed by various factors. For example, one may wish to employ primers from towards the termini of the total sequence.

Small polynucleotide segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly 10 practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR™ technology of U. S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular 15 biology.

The nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of the entire gene or gene fragments of interest. Depending on the application envisioned, one will typically desire to employ varying conditions of hybridization to achieve varying degrees of 20 selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, e.g., one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate 25 little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating related sequences.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template, less stringent (reduced stringency) hybridization conditions will typically be 30 needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ salt conditions such as those of from about 0.15 M to about 0.9 M

- salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to
5 destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

According to another embodiment of the present invention, polynucleotide compositions comprising antisense oligonucleotides are provided.
10 Antisense oligonucleotides have been demonstrated to be effective and targeted inhibitors of protein synthesis, and, consequently, provide a therapeutic approach by which a disease can be treated by inhibiting the synthesis of proteins that contribute to the disease. The efficacy of antisense oligonucleotides for inhibiting protein synthesis is well established. For example, the synthesis of polygalacturonase and the muscarine
15 type 2 acetylcholine receptor are inhibited by antisense oligonucleotides directed to their respective mRNA sequences (U. S. Patent 5,739,119 and U. S. Patent 5,759,829). Further, examples of antisense inhibition have been demonstrated with the nuclear protein cyclin, the multiple drug resistance gene (MDG1), ICAM-1, E-selectin, STK-1,
20 striatal GABA_A receptor and human EGF (Jaskulski *et al.*, Science. 1988 Jun 10;240(4858):1544-6; Vasantha Kumar and Ahmed, Cancer Commun. 1989;1(4):225-32; Peris *et al.*, Brain Res Mol Brain Res. 1998 Jun 15;57(2):310-20; U. S. Patent 5,801,154; U.S. Patent 5,789,573; U. S. Patent 5,718,709 and U.S. Patent 5,610,288). Antisense constructs have also been described that inhibit and can be used to treat a variety of abnormal cellular proliferations, *e.g.* cancer (U. S. Patent 5,747,470; U. S.
25 Patent 5,591,317 and U. S. Patent 5,783,683).

Therefore, in certain embodiments, the present invention provides oligonucleotide sequences that comprise all, or a portion of, any sequence that is capable of specifically binding to polynucleotide sequence described herein, or a complement thereof. In one embodiment, the antisense oligonucleotides comprise DNA
30 or derivatives thereof. In another embodiment, the oligonucleotides comprise RNA or derivatives thereof. In a third embodiment, the oligonucleotides are modified DNAs

comprising a phosphorothioated modified backbone. In a fourth embodiment, the oligonucleotide sequences comprise peptide nucleic acids or derivatives thereof. In each case, preferred compositions comprise a sequence region that is complementary, and more preferably substantially-complementary, and even more preferably, 5 completely complementary to one or more portions of polynucleotides disclosed herein. Selection of antisense compositions specific for a given gene sequence is based upon analysis of the chosen target sequence and determination of secondary structure, T_m , binding energy, and relative stability. Antisense compositions may be selected based upon their relative inability to form dimers, hairpins, or other secondary structures that 10 would reduce or prohibit specific binding to the target mRNA in a host cell. Highly preferred target regions of the mRNA, are those which are at or near the AUG translation initiation codon, and those sequences which are substantially complementary to 5' regions of the mRNA. These secondary structure analyses and target site selection considerations can be performed, for example, using v.4 of the OLIGO primer analysis 15 software and/or the BLASTN 2.0.5 algorithm software (Altschul *et al.*, Nucleic Acids Res. 1997, 25(17):3389-402).

The use of an antisense delivery method employing a short peptide vector, termed MPG (27 residues), is also contemplated. The MPG peptide contains a hydrophobic domain derived from the fusion sequence of HIV gp41 and a hydrophilic 20 domain from the nuclear localization sequence of SV40 T-antigen (Morris *et al.*, Nucleic Acids Res. 1997 Jul 15;25(14):2730-6). It has been demonstrated that several molecules of the MPG peptide coat the antisense oligonucleotides and can be delivered into cultured mammalian cells in less than 1 hour with relatively high efficiency (90%). Further, the interaction with MPG strongly increases both the stability of the 25 oligonucleotide to nuclease and the ability to cross the plasma membrane.

According to another embodiment of the invention, the polynucleotide compositions described herein are used in the design and preparation of ribozyme molecules for inhibiting expression of the tumor polypeptides and proteins of the present invention in tumor cells. Ribozymes are RNA-protein complexes that cleave 30 nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Cech, Proc Natl Acad Sci U S A. 1987

Dec;84(24):8788-92; Forster and Symons, Cell. 1987 Apr 24;49(2):211-20). For example, a large number of ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often cleaving only one of several phosphoesters in an oligonucleotide substrate (Cech *et al.*, Cell. 1981 Dec;27(3 Pt 2):487-96; Michel and Westhof, J Mol Biol. 1990 Dec 5;216(3):585-610; Reinhold-Hurek and Shub, Nature. 1992 May 14;357(6374):173-6). This specificity has been attributed to the requirement that the substrate bind via specific base-pairing interactions to the internal guide sequence ("IGS") of the ribozyme prior to chemical reaction.

Six basic varieties of naturally-occurring enzymatic RNAs are known presently. Each can catalyze the hydrolysis of RNA phosphodiester bonds *in trans* (and thus can cleave other RNA molecules) under physiological conditions. In general, enzymatic nucleic acids act by first binding to a target RNA. Such binding occurs through the target binding portion of a enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cut the target RNA. Strategic cleavage of such a target RNA will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic acid has bound and cleaved its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets.

The enzymatic nature of a ribozyme is advantageous over many technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its translation) since the concentration of ribozyme necessary to affect a therapeutic treatment is lower than that of an antisense oligonucleotide. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of target RNA. In addition, the ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding to the target RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site of cleavage can completely eliminate catalytic activity of a ribozyme. Similar mismatches in antisense molecules do not prevent their action

(Woolf *et al.*, Proc Natl Acad Sci U S A. 1992 Aug 15;89(16):7305-9). Thus, the specificity of action of a ribozyme is greater than that of an antisense oligonucleotide binding the same RNA site.

- The enzymatic nucleic acid molecule may be formed in a hammerhead, hairpin, a hepatitis δ virus, group I intron or RNaseP RNA (in association with an RNA guide sequence) or Neurospora VS RNA motif. Examples of hammerhead motifs are described by Rossi *et al.* Nucleic Acids Res. 1992 Sep 11;20(17):4559-65. Examples of hairpin motifs are described by Hampel *et al.* (Eur. Pat. Appl. Publ. No. EP 0360257), Hampel and Tritz, Biochemistry 1989 Jun 13;28(12):4929-33; Hampel *et al.*, Nucleic Acids Res. 1990 Jan 25;18(2):299-304 and U. S. Patent 5,631,359. An example of the hepatitis δ virus motif is described by Perrotta and Been, Biochemistry. 1992 Dec 1;31(47):11843-52; an example of the RNaseP motif is described by Guerrier-Takada *et al.*, Cell. 1983 Dec;35(3 Pt 2):849-57; Neurospora VS RNA ribozyme motif is described by Collins (Saville and Collins, Cell. 1990 May 18;61(4):685-96; Saville and Collins, Proc Natl Acad Sci U S A. 1991 Oct 1;88(19):8826-30; Collins and Olive, Biochemistry. 1993 Mar 23;32(11):2795-9); and an example of the Group I intron is described in (U. S. Patent 4,987,071). All that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or surrounding that substrate binding site which impart an RNA cleaving activity to the molecule. Thus the ribozyme constructs need not be limited to specific motifs mentioned herein.

Ribozymes may be designed as described in Int. Pat. Appl. Publ. No. WO 93/23569 and Int. Pat. Appl. Publ. No. WO 94/02595, each specifically incorporated herein by reference) and synthesized to be tested *in vitro* and *in vivo*, as described. Such ribozymes can also be optimized for delivery. While specific examples are provided, those in the art will recognize that equivalent RNA targets in other species can be utilized when necessary.

Ribozyme activity can be optimized by altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications that prevent their degradation by serum ribonucleases (see e.g., Int. Pat. Appl. Publ. No. WO

- 92/07065; Int. Pat. Appl. Publ. No. WO 93/15187; Int. Pat. Appl. Publ. No. WO 91/03162; Eur. Pat. Appl. Publ. No. 92110298.4; U. S. Patent 5,334,711; and Int. Pat. Appl. Publ. No. WO 94/13688, which describe various chemical modifications that can be made to the sugar moieties of enzymatic RNA molecules), modifications which
5 enhance their efficacy in cells, and removal of stem II bases to shorten RNA synthesis times and reduce chemical requirements.

Sullivan *et al.* (Int. Pat. Appl. Publ. No. WO 94/02595) describes the general methods for delivery of enzymatic RNA molecules. Ribozymes may be administered to cells by a variety of methods known to those familiar to the art,
10 including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, ribozymes may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the RNA/vehicle combination may be locally delivered by direct
15 inhalation, by direct injection or by use of a catheter, infusion pump or stent. Other routes of delivery include, but are not limited to, intravascular, intramuscular, subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions of ribozyme delivery and administration are provided in Int. Pat. Appl. Publ. No. WO
20 94/02595 and Int. Pat. Appl. Publ. No. WO 93/23569, each specifically incorporated herein by reference.

Another means of accumulating high concentrations of a ribozyme(s) within cells is to incorporate the ribozyme-encoding sequences into a DNA expression vector. Transcription of the ribozyme sequences are driven from a promoter for
25 eukaryotic RNA polymerase I (pol I), RNA polymerase II (pol II), or RNA polymerase III (pol III). Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, *etc.*) present nearby. Prokaryotic RNA polymerase promoters may also be used, providing that the
30 prokaryotic RNA polymerase enzyme is expressed in the appropriate cells. Ribozymes expressed from such promoters have been shown to function in mammalian cells. Such

transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-associated vectors), or viral RNA vectors (such as retroviral, semliki forest virus, sindbis virus vectors).

5 In another embodiment of the invention, peptide nucleic acids (PNAs) compositions are provided. PNA is a DNA mimic in which the nucleobases are attached to a pseudopeptide backbone (Good and Nielsen, *Antisense Nucleic Acid Drug Dev.* 1997 7(4) 431-37). PNA is able to be utilized in a number methods that traditionally have used RNA or DNA. Often PNA sequences perform better in
10 techniques than the corresponding RNA or DNA sequences and have utilities that are not inherent to RNA or DNA. A review of PNA including methods of making, characteristics of, and methods of using, is provided by Corey (*Trends Biotechnol* 1997 Jun;15(6):224-9). As such, in certain embodiments, one may prepare PNA sequences that are complementary to one or more portions of the ACE mRNA sequence, and such
15 PNA compositions may be used to regulate, alter, decrease, or reduce the translation of ACE-specific mRNA, and thereby alter the level of ACE activity in a host cell to which such PNA compositions have been administered.

PNAs have 2-aminoethyl-glycine linkages replacing the normal phosphodiester backbone of DNA (Nielsen *et al.*, *Science* 1991 Dec 6;254(5037):1497-
20 500; Hanvey *et al.*, *Science*. 1992 Nov 27;258(5087):1481-5; Hyrup and Nielsen, *Bioorg Med Chem.* 1996 Jan;4(1):5-23). This chemistry has three important consequences: firstly, in contrast to DNA or phosphorothioate oligonucleotides, PNAs are neutral molecules; secondly, PNAs are achiral, which avoids the need to develop a stereoselective synthesis; and thirdly, PNA synthesis uses standard Boc or Fmoc
25 protocols for solid-phase peptide synthesis, although other methods, including a modified Merrifield method, have been used.

PNA monomers or ready-made oligomers are commercially available from PerSeptive Biosystems (Framingham, MA). PNA syntheses by either Boc or Fmoc protocols are straightforward using manual or automated protocols (Norton *et al.*,
30 *Bioorg Med Chem.* 1995 Apr;3(4):437-45). The manual protocol lends itself to the

production of chemically modified PNAs or the simultaneous synthesis of families of closely related PNAs.

- As with peptide synthesis, the success of a particular PNA synthesis will depend on the properties of the chosen sequence. For example, while in theory PNAs can incorporate any combination of nucleotide bases, the presence of adjacent purines can lead to deletions of one or more residues in the product. In expectation of this difficulty, it is suggested that, in producing PNAs with adjacent purines, one should repeat the coupling of residues likely to be added inefficiently. This should be followed by the purification of PNAs by reverse-phase high-pressure liquid chromatography, providing yields and purity of product similar to those observed during the synthesis of peptides.

- Modifications of PNAs for a given application may be accomplished by coupling amino acids during solid-phase synthesis or by attaching compounds that contain a carboxylic acid group to the exposed N-terminal amine. Alternatively, PNAs can be modified after synthesis by coupling to an introduced lysine or cysteine. The ease with which PNAs can be modified facilitates optimization for better solubility or for specific functional requirements. Once synthesized, the identity of PNAs and their derivatives can be confirmed by mass spectrometry. Several studies have made and utilized modifications of PNAs (for example, Norton *et al.*, Bioorg Med Chem. 1995 Apr;3(4):437-45; Petersen *et al.*, J Pept Sci. 1995 May-Jun;1(3):175-83; Orum *et al.*, Biotechniques. 1995 Sep;19(3):472-80; Footer *et al.*, Biochemistry. 1996 Aug 20;35(33):10673-9; Griffith *et al.*, Nucleic Acids Res. 1995 Aug 11;23(15):3003-8; Pardridge *et al.*, Proc Natl Acad Sci U S A. 1995 Jun 6;92(12):5592-6; Boffa *et al.*, Proc Natl Acad Sci U S A. 1995 Mar 14;92(6):1901-5; Gambacorti-Passerini *et al.*, Blood. 1996 Aug 15;88(4):1411-7; Armitage *et al.*, Proc Natl Acad Sci U S A. 1997 Nov 11;94(23):12320-5; Seeger *et al.*, Biotechniques. 1997 Sep;23(3):512-7). U.S. Patent No. 5,700,922 discusses PNA-DNA-PNA chimeric molecules and their uses in diagnostics, modulating protein in organisms, and treatment of conditions susceptible to therapeutics.
- Methods of characterizing the antisense binding properties of PNAs are discussed in Rose (Anal Chem. 1993 Dec 15;65(24):3545-9) and Jensen *et al.*

(*Biochemistry*. 1997 Apr 22;36(16):5072-7). Rose uses capillary gel electrophoresis to determine binding of PNAs to their complementary oligonucleotide, measuring the relative binding kinetics and stoichiometry. Similar types of measurements were made by Jensen *et al.* using BIAcore™ technology.

5 Other applications of PNAs that have been described and will be apparent to the skilled artisan include use in DNA strand invasion, antisense inhibition, mutational analysis, enhancers of transcription, nucleic acid purification, isolation of transcriptionally active genes, blocking of transcription factor binding, genome cleavage, biosensors, *in situ* hybridization, and the like.

10 Polynucleotide Identification, Characterization and Expression

Polynucleotides compositions of the present invention may be identified, prepared and/or manipulated using any of a variety of well established techniques (see generally, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989, and other like references). For 15 example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.*, expression that is at least two fold greater in a tumor than in normal tissue, as determined using a representative assay provided herein). Such screens may be performed, for example, using the microarray technology of Affymetrix, Inc. (Santa Clara, CA) according to the 20 manufacturer's instructions (and essentially as described by Schena et al., *Proc. Natl. Acad. Sci. USA* 93:10614-10619, 1996 and Heller et al., *Proc. Natl. Acad. Sci. USA* 94:2150-2155, 1997). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as tumor cells.

Many template dependent processes are available to amplify a target 25 sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR™) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in its entirety. Briefly, in PCR™, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target 30 sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture

along with a DNA polymerase (*e.g.*, *Taq* polymerase). If the target sequence is present in a sample, the primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will 5 dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCR™ amplification procedure may be performed in order to quantify the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

10 Any of a number of other template dependent processes, many of which are variations of the PCR™ amplification technique, are readily known and available in the art. Illustratively, some such methods include the ligase chain reaction (referred to as LCR), described, for example, in Eur. Pat. Appl. Publ. No. 320,308 and U.S. Patent No. 4,883,750; Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. 15 PCT/US87/00880; Strand Displacement Amplification (SDA) and Repair Chain Reaction (RCR). Still other amplification methods are described in Great Britain Pat. Appl. No. 2 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025. Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (PCT Intl. Pat. Appl. Publ. No. WO 88/10315), including nucleic acid sequence 20 based amplification (NASBA) and 3SR. Eur. Pat. Appl. Publ. No. 329,822 describes a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA). PCT Intl. Pat. Appl. Publ. No. WO 89/06700 describes a nucleic acid sequence amplification scheme based 25 on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. Other amplification methods such as "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) are also well-known to those of skill in the art.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (*e.g.*, a tumor cDNA library) 30 using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification.

Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

- For hybridization techniques, a partial sequence may be labeled (e.g., by
- 5 nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (see Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are
 - 10 selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may
 - 15 involve generating a series of deletion clones. The resulting overlapping sequences can then assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

- Alternatively, amplification techniques, such as those described above, can be useful for obtaining a full length coding sequence from a partial cDNA sequence.
- 20 One such amplification technique is inverse PCR (see Triglia et al., *Nucl. Acids Res.* 16:8186, 1988), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be
 - 25 retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO
 - 30 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer,

which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., *PCR Methods Appl.* 1:111-19, 1991) and walking PCR (Parker et al., *Nucl. Acids. Res.* 19:3055-60, 1991). Other methods employing amplification may also be employed 5 to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (e.g., NCBI BLAST searches), and such ESTs 10 may be used to generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct 15 expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous 20 in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring 25 sequence.

Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For 30 example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide

sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

- In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

- Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 215-223, Horn, T. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) *Science* 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

- A newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) *Proteins, Structures and Molecular Principles*, WH Freeman and Co., New York, N.Y.) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

- In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the

transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described, for example, in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

10 A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell 15 systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 20 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid 25 lacZ promoter of the pBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or pSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be 30 advantageously used with an appropriate selectable marker.

In bacterial systems, any of a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as pBLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of .beta.-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) *Methods Enzymol.* 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311. Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or

Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

- An insect system may also be used to express a polypeptide of interest. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus 5 (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat 10 protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which the polypeptide of interest may be expressed (Engelhard, E. K. et al. (1994) *Proc. Natl. Acad. Sci.* 91 :3224-3227).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression 15 vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, 20 transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the 25 polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct 30 reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic.

The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

- In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "pro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as 10 CHO, COS, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

- For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a 15 polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer 20 resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

- Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine 25 kinase (Wigler, M. et al. (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1990) *Cell* 22:817-23) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); npt, which 30 confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150:1-14); and als or pat, which confer resistance to

chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histidol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-51). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) *Methods Mol. Biol.* 55:121-131).

10 Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem
15 with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

20 Alternatively, host cells that contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include, for example, membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

25 A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be
30 preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990;

Serological Methods, a Laboratory Manual, APS Press, St Paul. Minn.) and Maddox, D. E. et al. (1983; *J. Exp. Med.* 158:1211-1216).

- A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means
- 5 for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA
- 10 probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

- 15 Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the
- 20 invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are
- 25 not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego,
- 30 Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion

protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath, J. et al. (1992, *Prot. Exp. Purif.* 3:263-281) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; *DNA Cell Biol.* 12:441-453).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) *J. Am. Chem. Soc.* 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

Antibody Compositions, Fragments Thereof and Other Binding Agents

According to another aspect, the present invention further provides binding agents, such as antibodies and antigen-binding fragments thereof, that exhibit immunological binding to a tumor polypeptide disclosed herein, or to a portion, variant or derivative thereof. An antibody, or antigen-binding fragment thereof, is said to "specifically bind," "immunologically bind," and/or is "immunologically reactive" to a polypeptide of the invention if it reacts at a detectable level (within, for example, an ELISA assay) with the polypeptide, and does not react detectably with unrelated polypeptides under similar conditions.

Immunological binding, as used in this context, generally refers to the non-covalent interactions of the type which occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific. The strength, or affinity of immunological binding interactions can be expressed in terms of the dissociation constant (K_d) of the interaction, wherein a smaller K_d represents a greater affinity. Immunological binding properties of selected polypeptides can be quantified

using methods well known in the art. One such method entails measuring the rates of antigen-binding site/antigen complex formation and dissociation, wherein those rates depend on the concentrations of the complex partners, the affinity of the interaction, and on geometric parameters that equally influence the rate in both directions. Thus, both 5 the "on rate constant" (K_{on}) and the "off rate constant" (K_{off}) can be determined by calculation of the concentrations and the actual rates of association and dissociation. The ratio of K_{off}/K_{on} enables cancellation of all parameters not related to affinity, and is thus equal to the dissociation constant K_d . See, generally, Davies et al. (1990) Annual Rev. Biochem. 59:439-473.

10 An "antigen-binding site," or "binding portion" of an antibody refers to the part of the immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable ("V") regions of the heavy ("H") and light ("L") chains. Three highly divergent stretches within the V regions of the heavy and light chains are referred to as "hypervariable 15 regions" which are interposed between more conserved flanking stretches known as "framework regions," or "FRs". Thus the term "FR" refers to amino acid sequences which are naturally found between and adjacent to hypervariable regions in immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each 20 other in three dimensional space to form an antigen-binding surface. The antigen-binding surface is complementary to the three-dimensional surface of a bound antigen, and the three hypervariable regions of each of the heavy and light chains are referred to as "complementarity-determining regions," or "CDRs."

Binding agents may be further capable of differentiating between patients 25 with and without a cancer, such as lung cancer, using the representative assays provided herein. For example, antibodies or other binding agents that bind to a tumor protein will preferably generate a signal indicating the presence of a cancer in at least about 20% of patients with the disease, more preferably at least about 30% of patients. Alternatively, or in addition, the antibody will generate a negative signal indicating the 30 absence of the disease in at least about 90% of individuals without the cancer. To determine whether a binding agent satisfies this requirement, biological samples (e.g.,

- blood, sera, sputum, urine and/or tumor biopsies) from patients with and without a cancer (as determined using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. Preferably, a statistically significant number of samples with and without the disease will be assayed. Each 5 binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may be used in combination to improve sensitivity.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, 10 an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. *See, e.g., Harlow and Lane, Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, antibodies can be produced by cell culture techniques, including the generation 15 of monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (*e.g.*, mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen 20 without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. 25 Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. 30 Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the

desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized 5 animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, 10 colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the 15 yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process 20 in, for example, an affinity chromatography step.

A number of therapeutically useful molecules are known in the art which comprise antigen-binding sites that are capable of exhibiting immunological binding properties of an antibody molecule. The proteolytic enzyme papain preferentially cleaves IgG molecules to yield several fragments, two of which (the "F(ab)" fragments) 25 each comprise a covalent heterodimer that includes an intact antigen-binding site. The enzyme pepsin is able to cleave IgG molecules to provide several fragments, including the "F(ab')₂" fragment which comprises both antigen-binding sites. An "Fv" fragment can be produced by preferential proteolytic cleavage of an IgM, and on rare occasions IgG or IgA immunoglobulin molecule. Fv fragments are, however, more commonly 30 derived using recombinant techniques known in the art. The Fv fragment includes a non-covalent V_H::V_L heterodimer including an antigen-binding site which retains much

of the antigen recognition and binding capabilities of the native antibody molecule. Inbar et al. (1972) Proc. Nat. Acad. Sci. USA 69:2659-2662; Hochman et al. (1976) Biochem 15:2706-2710; and Ehrlich et al. (1980) Biochem 19:4091-4096.

- A single chain Fv ("sFv") polypeptide is a covalently linked $V_H::V_L$ heterodimer which is expressed from a gene fusion including V_H - and V_L -encoding genes linked by a peptide-encoding linker. Huston et al. (1988) Proc. Nat. Acad. Sci. USA 85(16):5879-5883. A number of methods have been described to discern chemical structures for converting the naturally aggregated--but chemically separated--light and heavy polypeptide chains from an antibody V region into an sFv molecule which will 10 fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, e.g., U.S. Pat. Nos. 5,091,513 and 5,132,405, to Huston et al.; and U.S. Pat. No. 4,946,778, to Ladner et al.

Each of the above-described molecules includes a heavy chain and a light chain CDR set, respectively interposed between a heavy chain and a light chain FR set which provide support to the CDRS and define the spatial relationship of the CDRs relative to each other. As used herein, the term "CDR set" refers to the three hypervariable regions of a heavy or light chain V region. Proceeding from the N-terminus of a heavy or light chain, these regions are denoted as "CDR1," "CDR2," and "CDR3" respectively. An antigen-binding site, therefore, includes six CDRs, 20 comprising the CDR set from each of a heavy and a light chain V region. A polypeptide comprising a single CDR, (e.g., a CDR1, CDR2 or CDR3) is referred to herein as a "molecular recognition unit." Crystallographic analysis of a number of antigen-antibody complexes has demonstrated that the amino acid residues of CDRs form extensive contact with bound antigen, wherein the most extensive antigen contact is with the 25 heavy chain CDR3. Thus, the molecular recognition units are primarily responsible for the specificity of an antigen-binding site.

As used herein, the term "FR set" refers to the four flanking amino acid sequences which frame the CDRs of a CDR set of a heavy or light chain V region. Some FR residues may contact bound antigen; however, FRs are primarily responsible 30 for folding the V region into the antigen-binding site, particularly the FR residues directly adjacent to the CDRS. Within FRs, certain amino residues and certain structural

features are very highly conserved. In this regard, all V region sequences contain an internal disulfide loop of around 90 amino acid residues. When the V regions fold into a binding-site, the CDRs are displayed as projecting loop motifs which form an antigen-binding surface. It is generally recognized that there are conserved structural regions of
5 FRs which influence the folded shape of the CDR loops into certain "canonical" structures--regardless of the precise CDR amino acid sequence. Further, certain FR residues are known to participate in non-covalent interdomain contacts which stabilize the interaction of the antibody heavy and light chains.

A number of "humanized" antibody molecules comprising an antigen-
10 binding site derived from a non-human immunoglobulin have been described, including chimeric antibodies having rodent V regions and their associated CDRs fused to human constant domains (Winter et al. (1991) Nature 349:293-299; Lobuglio et al. (1989) Proc. Nat. Acad. Sci. USA 86:4220-4224; Shaw et al. (1987) J Immunol. 138:4534-4538; and Brown et al. (1987) Cancer Res. 47:3577-3583), rodent CDRs grafted into a
15 human supporting FR prior to fusion with an appropriate human antibody constant domain (Riechmann et al. (1988) Nature 332:323-327; Verhoeyen et al. (1988) Science 239:1534-1536; and Jones et al. (1986) Nature 321:522-525), and rodent CDRs supported by recombinantly veneered rodent FRs (European Patent Publication No. 519,596, published Dec. 23, 1992). These "humanized" molecules are designed to
20 minimize unwanted immunological response toward rodent antihuman antibody molecules which limits the duration and effectiveness of therapeutic applications of those moieties in human recipients.

As used herein, the terms "veneered FRs" and "recombinantly veneered FRs" refer to the selective replacement of FR residues from, e.g., a rodent heavy or light
25 chain V region, with human FR residues in order to provide a xenogeneic molecule comprising an antigen-binding site which retains substantially all of the native FR polypeptide folding structure. Veneering techniques are based on the understanding that the ligand binding characteristics of an antigen-binding site are determined primarily by the structure and relative disposition of the heavy and light chain CDR sets within the
30 antigen-binding surface. Davies et al. (1990) Ann. Rev. Biochem. 59:439-473. Thus, antigen binding specificity can be preserved in a humanized antibody only wherein the

- CDR structures, their interaction with each other, and their interaction with the rest of the V region domains are carefully maintained. By using veneering techniques, exterior (e.g., solvent-accessible) FR residues which are readily encountered by the immune system are selectively replaced with human residues to provide a hybrid molecule that
- 5 comprises either a weakly immunogenic, or substantially non-immunogenic veneered surface.

The process of veneering makes use of the available sequence data for human antibody variable domains compiled by Kabat et al., in Sequences of Proteins of Immunological Interest, 4th ed., (U.S. Dept. of Health and Human Services, U.S.

10 Government Printing Office, 1987), updates to the Kabat database, and other accessible U.S. and foreign databases (both nucleic acid and protein). Solvent accessibilities of V region amino acids can be deduced from the known three-dimensional structure for human and murine antibody fragments. There are two general steps in veneering a murine antigen-binding site. Initially, the FRs of the variable domains of an antibody

15 molecule of interest are compared with corresponding FR sequences of human variable domains obtained from the above-identified sources. The most homologous human V regions are then compared residue by residue to corresponding murine amino acids. The residues in the murine FR which differ from the human counterpart are replaced by the residues present in the human moiety using recombinant techniques well known in the

20 art. Residue switching is only carried out with moieties which are at least partially exposed (solvent accessible), and care is exercised in the replacement of amino acid residues which may have a significant effect on the tertiary structure of V region domains, such as proline, glycine and charged amino acids.

In this manner, the resultant "veneered" murine antigen-binding sites are

25 thus designed to retain the murine CDR residues, the residues substantially adjacent to the CDRs, the residues identified as buried or mostly buried (solvent inaccessible), the residues believed to participate in non-covalent (e.g., electrostatic and hydrophobic) contacts between heavy and light chain domains, and the residues from conserved structural regions of the FRs which are believed to influence the "canonical" tertiary

30 structures of the CDR loops. These design criteria are then used to prepare recombinant nucleotide sequences which combine the CDRs of both the heavy and light chain of a

murine antigen-binding site into human-appearing FRs that can be used to transfect mammalian cells for the expression of recombinant human antibodies which exhibit the antigen specificity of the murine antibody molecule.

In another embodiment of the invention, monoclonal antibodies of the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include ⁹⁰Y, ¹²³I, ¹²⁵I, ¹³¹I, ¹⁸⁶Re, ¹⁸⁸Re, ²¹¹At, and ²¹²Bi. Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diphtheria toxin, cholera toxin, gelonin, *Pseudomonas* exotoxin, *Shigella* toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a suitable monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (*e.g.*, a halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, *e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.

Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the 5 intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710, to Spitzer), by irradiation of a photolabile bond (*e.g.*, U.S. Patent No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains (*e.g.*, U.S. Patent No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (*e.g.*, U.S. Patent No. 4,671,958, to Rodwell 10 et al.), and acid-catalyzed hydrolysis (*e.g.*, U.S. Patent No. 4,569,789, to Blattler et al.).

It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent 15 may be prepared in a variety of ways. For example, more than one agent may be coupled directly to an antibody molecule, or linkers that provide multiple sites for attachment can be used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as 20 albumins (*e.g.*, U.S. Patent No. 4,507,234, to Kato et al.), peptides and polysaccharides such as aminodextran (*e.g.*, U.S. Patent No. 4,699,784, to Shih et al.). A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (*e.g.*, U.S. Patent Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating 25 compounds. For example, U.S. Patent No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison et al. discloses representative chelating 30 compounds and their synthesis.

T Cell Compositions

- The present invention, in another aspect, provides T cells specific for a tumor polypeptide disclosed herein, or for a variant or derivative thereof. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example,
- 5 T cells may be isolated from bone marrow, peripheral blood, or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as the Isolex™ System, available from Nexell Therapeutics, Inc. (Irvine, CA; see also U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or
- 10 unrelated humans, non-human mammals, cell lines or cultures.

T cells may be stimulated with a polypeptide, polynucleotide encoding a polypeptide and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide of interest. Preferably, a tumor

15 polypeptide or polynucleotide of the invention is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

T cells are considered to be specific for a polypeptide of the present invention if the T cells specifically proliferate, secrete cytokines or kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell

20 specificity may be evaluated using any of a variety of standard techniques. For example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen et al., *Cancer Res.* 54:1065-1070, 1994. Alternatively, detection of the

25 proliferation of T cells may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (*e.g.*, by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a tumor polypeptide (100 ng/ml - 100 µg/ml, preferably 200 ng/ml - 25 µg/ml) for 3 - 7

30 days will typically result in at least a two fold increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as

measured using standard cytokine assays in which a two fold increase in the level of cytokine release (*e.g.*, TNF or IFN- γ) is indicative of T cell activation (*see* Coligan et al., Current Protocols in Immunology, vol. 1, Wiley Interscience (Greene 1998)). T cells that have been activated in response to a tumor polypeptide, polynucleotide or 5 polypeptide-expressing APC may be CD4 $^{+}$ and/or CD8 $^{+}$. Tumor polypeptide-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T cells are derived from a patient, a related donor or an unrelated donor, and are administered to the patient following stimulation and expansion.

For therapeutic purposes, CD4 $^{+}$ or CD8 $^{+}$ T cells that proliferate in 10 response to a tumor polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example, the T cells can be re-exposed to a tumor polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator 15 cells that synthesize a tumor polypeptide. Alternatively, one or more T cells that proliferate in the presence of the tumor polypeptide can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

T Cell Receptor Compositions

20 The T cell receptor (TCR) consists of 2 different, highly variable polypeptide chains, termed the T-cell receptor α and β chains, that are linked by a disulfide bond (Janeway, Travers, Walport. *Immunobiology*. Fourth Ed., 148-159. Elsevier Science Ltd/Garland Publishing. 1999). The α/β heterodimer complexes with the invariant CD3 chains at the cell membrane. This complex recognizes specific 25 antigenic peptides bound to MHC molecules. The enormous diversity of TCR specificities is generated much like immunoglobulin diversity, through somatic gene rearrangement. The β chain genes contain over 50 variable (V), 2 diversity (D), over 10 joining (J) segments, and 2 constant region segments (C). The α chain genes contain over 70 V segments, and over 60 J segments but no D segments, as well as one C 30 segment. During T cell development in the thymus, the D to J gene rearrangement of

the β chain occurs, followed by the V gene segment rearrangement to the DJ. This functional VDJ β exon is transcribed and spliced to join to a C β . For the α chain, a V α gene segment rearranges to a J α gene segment to create the functional exon that is then transcribed and spliced to the C α . Diversity is further increased during the 5 recombination process by the random addition of P and N-nucleotides between the V, D, and J segments of the β chain and between the V and J segments in the α chain (Janeway, Travers, Walport. *Immunobiology*. Fourth Ed., 98 and 150. Elsevier Science Ltd/Garland Publishing, 1999).

The present invention, in another aspect, provides TCRs specific for a 10 polypeptide disclosed herein, or for a variant or derivative thereof. In accordance with the present invention, polynucleotide and amino acid sequences are provided for the V-J or V-D-J junctional regions or parts thereof for the alpha and beta chains of the T-cell receptor which recognize tumor polypeptides described herein. In general, this aspect of the invention relates to T-cell receptors which recognize or bind tumor polypeptides 15 presented in the context of MHC. In a preferred embodiment the tumor antigens recognized by the T-cell receptors comprise a polypeptide of the present invention. For example, cDNA encoding a TCR specific for a _tumor peptide can be isolated from T cells specific for a tumor polypeptide using standard molecular biological and recombinant DNA techniques.

20 This invention further includes the T-cell receptors or analogs thereof having substantially the same function or activity as the T-cell receptors of this invention which recognize or bind tumor polypeptides. Such receptors include, but are not limited to, a fragment of the receptor, or a substitution, addition or deletion mutant of a T-cell receptor provided herein. This invention also encompasses polypeptides or 25 peptides that are substantially homologous to the T-cell receptors provided herein or that retain substantially the same activity. The term "analog" includes any protein or polypeptide having an amino acid residue sequence substantially identical to the T-cell receptors provided herein in which one or more residues, preferably no more than 5 residues, more preferably no more than 25 residues have been conservatively substituted 30 with a functionally similar residue and which displays the functional aspects of the T-cell receptor as described herein.

The present invention further provides for suitable mammalian host cells, for example, non-specific T cells, that are transfected with a polynucleotide encoding TCRs specific for a polypeptide described herein, thereby rendering the host cell specific for the polypeptide. The α and β chains of the TCR may be contained on 5 separate expression vectors or alternatively, on a single expression vector that also contains an internal ribosome entry site (IRES) for cap-independent translation of the gene downstream of the IRES. Said host cells expressing TCRs specific for the polypeptide may be used, for example, for adoptive immunotherapy of lung cancer as discussed further below.

10 In further aspects of the present invention, cloned TCRs specific for a polypeptide recited herein may be used in a kit for the diagnosis of lung cancer. For example, the nucleic acid sequence or portions thereof, of tumor-specific TCRs can be used as probes or primers for the detection of expression of the rearranged genes encoding the specific TCR in a biological sample. Therefore, the present invention 15 further provides for an assay for detecting messenger RNA or DNA encoding the TCR specific for a polypeptide.

Pharmaceutical Compositions

In additional embodiments, the present invention concerns formulation of one or more of the polynucleotide, polypeptide, T-cell, TCR, and/or antibody 20 compositions disclosed herein in pharmaceutically-acceptable carriers for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

It will be understood that, if desired, a composition as disclosed herein may be administered in combination with other agents as well, such as, e.g., other 25 proteins or polypeptides or various pharmaceutically-active agents. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the particular instance. Such compositions may be purified from

host cells or other biological sources, or alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

Therefore, in another aspect of the present invention, pharmaceutical compositions are provided comprising one or more of the polynucleotide, polypeptide, antibody, TCR, and/or T-cell compositions described herein in combination with a physiologically acceptable carrier. In certain preferred embodiments, the pharmaceutical compositions of the invention comprise immunogenic polynucleotide and/or polypeptide compositions of the invention for use in prophylactic and therapeutic vaccine applications. Vaccine preparation is generally described in, for example, M.F. Powell and M.J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995). Generally, such compositions will comprise one or more polynucleotide and/or polypeptide compositions of the present invention in combination with one or more immunostimulants.

It will be apparent that any of the pharmaceutical compositions described herein can contain pharmaceutically acceptable salts of the polynucleotides and polypeptides of the invention. Such salts can be prepared, for example, from pharmaceutically acceptable non-toxic bases, including organic bases (e.g., salts of primary, secondary and tertiary amines and basic amino acids) and inorganic bases (e.g., sodium, potassium, lithium, ammonium, calcium and magnesium salts).

In another embodiment, illustrative immunogenic compositions, e.g., vaccine compositions, of the present invention comprise DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the polynucleotide may be administered within any of a variety of delivery systems known to those of ordinary skill in the art. Indeed, numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems* 15:143-198, 1998, and references cited therein. Appropriate polynucleotide expression systems will, of course, contain the necessary regulatory DNA regulatory sequences for expression in a patient (such as a suitable promoter and terminating signal). Alternatively, bacterial delivery systems may involve

the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope.

Therefore, in certain embodiments, polynucleotides encoding immunogenic polypeptides described herein are introduced into suitable mammalian host cells for expression using any of a number of known viral-based systems. In one illustrative embodiment, retroviruses provide a convenient and effective platform for gene delivery systems. A selected nucleotide sequence encoding a polypeptide of the present invention can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to a subject. A number of illustrative retroviral systems have been described (e.g., U.S. Pat. No. 5,219,740; Miller and Rosman (1989) BioTechniques 7:980-990; Miller, A. D. (1990) Human Gene Therapy 1:5-14; Scarpa et al. (1991) Virology 180:849-852; Burns et al. (1993) Proc. Natl. Acad. Sci. USA 90:8033-8037; and Boris-Lawrie and Temin (1993) Cur. Opin. Genet. Develop. 3:102-109.

In addition, a number of illustrative adenovirus-based systems have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham (1986) J. Virol. 57:267-274; Bett et al. (1993) J. Virol. 67:5911-5921; Mittereder et al. (1994) Human Gene Therapy 5:717-729; Seth et al. (1994) J. Virol. 68:933-940; Barr et al. (1994) Gene Therapy 1:51-58; Berkner, K. L. (1988) BioTechniques 6:616-629; and Rich et al. (1993) Human Gene Therapy 4:461-476).

Various adeno-associated virus (AAV) vector systems have also been developed for polynucleotide delivery. AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Pat. Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 and WO 93/03769; Lebkowski et al. (1988) Molec. Cell. Biol. 8:3988-3996; Vincent et al. (1990) Vaccines 90 (Cold Spring Harbor Laboratory Press); Carter, B. J. (1992) Current Opinion in Biotechnology 3:533-539; Muzyczka, N. (1992) Current Topics in Microbiol. and Immunol. 158:97-129; Kotin, R. M. (1994) Human Gene Therapy 5:793-801; Shelling and Smith (1994) Gene Therapy 1:165-169; and Zhou et al. (1994) J. Exp. Med. 179:1867-1875.

- Additional viral vectors useful for delivering the polynucleotides encoding polypeptides of the present invention by gene transfer include those derived from the pox family of viruses, such as vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the novel molecules can be
- 5 constructed as follows. The DNA encoding a polypeptide is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene
- 10 encoding the polypeptide of interest into the viral genome. The resulting TK.sup.(-) recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

A vaccinia-based infection/transfection system can be conveniently used to provide for inducible, transient expression or coexpression of one or more

15 polypeptides described herein in host cells of an organism. In this particular system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide or polynucleotides of interest, driven by a T7

20 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into polypeptide by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, e.g., Elroy-Stein and Moss, Proc. Natl. Acad. Sci. USA (1990) 87:6743-

25 6747; Fuerst et al. Proc. Natl. Acad. Sci. USA (1986) 83:8122-8126.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the coding sequences of interest. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an Avipox

30 vector is particularly desirable in human and other mammalian species since members of the Avipox genus can only productively replicate in susceptible avian species and

therefore are not infective in mammalian cells. Methods for producing recombinant Avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

5 Any of a number of alphavirus vectors can also be used for delivery of polynucleotide compositions of the present invention, such as those vectors described in U.S. Patent Nos. 5,843,723; 6,015,686; 6,008,035 and 6,015,694. Certain vectors based on Venezuelan Equine Encephalitis (VEE) can also be used, illustrative examples of which can be found in U.S. Patent Nos. 5,505,947 and 5,643,576.

10 Moreover, molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al. *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al. *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery under the invention.

Additional illustrative information on these and other known viral-based
15 delivery systems can be found, for example, in Fisher-Hoch et al., *Proc. Natl. Acad. Sci. USA* 86:317-321, 1989; Flexner et al., *Ann. N.Y. Acad. Sci.* 569:86-103, 1989; Flexner et al., *Vaccine* 8:17-21, 1990; U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques* 6:616-627, 1988; Rosenfeld et al., *Science* 252:431-434, 1991; Kolls et al., *Proc. Natl. Acad. Sci. USA* 91:215-219, 1994; Kass-Eisler et al., *Proc. Natl. Acad. Sci. USA* 90:11498-11502, 1993; Guzman et al., *Circulation* 88:2838-2848, 1993; and Guzman et al., *Cir. Res.* 73:1202-1207, 1993.

In certain embodiments, a polynucleotide may be integrated into the genome of a target cell. This integration may be in the specific location and orientation
25 via homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the polynucleotide may be stably maintained in the cell as a separate, episomal segment of DNA. Such polynucleotide segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host
30 cell cycle. The manner in which the expression construct is delivered to a cell and

where in the cell the polynucleotide remains is dependent on the type of expression construct employed.

In another embodiment of the invention, a polynucleotide is administered/delivered as "naked" DNA, for example as described in Ulmer et al., 5 *Science* 259:1745-1749, 1993 and reviewed by Cohen, *Science* 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

In still another embodiment, a composition of the present invention can be delivered via a particle bombardment approach, many of which have been described. 10 In one illustrative example, gas-driven particle acceleration can be achieved with devices such as those manufactured by Powderject Pharmaceuticals PLC (Oxford, UK) and Powderject Vaccines Inc. (Madison, WI), some examples of which are described in U.S. Patent Nos. 5,846,796; 6,010,478; 5,865,796; 5,584,807; and EP Patent No. 0500 799. This approach offers a needle-free delivery approach wherein a dry powder 15 formulation of microscopic particles, such as polynucleotide or polypeptide particles, are accelerated to high speed within a helium gas jet generated by a hand held device, propelling the particles into a target tissue of interest.

In a related embodiment, other devices and methods that may be useful for gas-driven needle-less injection of compositions of the present invention include 20 those provided by Bioject, Inc. (Portland, OR), some examples of which are described in U.S. Patent Nos. 4,790,824; 5,064,413; 5,312,335; 5,383,851; 5,399,163; 5,520,639 and 5,993,412.

According to another embodiment, the pharmaceutical compositions described herein will comprise one or more immunostimulants in addition to the 25 immunogenic polynucleotide, polypeptide, antibody, T-cell, TCR, and/or APC compositions of this invention. An immunostimulant refers to essentially any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen. One preferred type of immunostimulant comprises an adjuvant. Many adjuvants contain a substance designed to protect the antigen from rapid 30 catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bordetella pertussis* or *Mycobacterium tuberculosis* derived

- proteins. Certain adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quill A. Cytokines, such as GM-CSF, interleukin-2, -7, -12, and other like growth factors, may also be used as adjuvants.

Within certain embodiments of the invention, the adjuvant composition is preferably one that induces an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (*e.g.*, IFN- γ , TNF α , IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (*e.g.*, IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol.* 7:145-173, 1989.

Certain preferred adjuvants for eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A, together with an aluminum salt. MPL® adjuvants are available from Corixa Corporation (Seattle, WA; *see*, for example, US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato et al., *Science* 273:352, 1996. Another preferred adjuvant comprises a saponin,

such as Quil A, or derivatives thereof, including QS21 and QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or *Gypsophila* or *Chenopodium quinoa* saponins. Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example 5 combinations of at least two of the following group comprising QS21, QS7, Quil A, β -escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix, 10 particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs. Furthermore, the saponins may be formulated together with a polyoxyethylene ether or ester, in either a non-particulate solution or 15 suspension, or in a particulate structure such as a paucilamellar liposome or ISCOM. The saponins may also be formulated with excipients such as Carbopol^R to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.

In one preferred embodiment, the adjuvant system includes the 20 combination of a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL[®] adjuvant, as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. Another particularly preferred adjuvant formulation employing QS21, 3D- 25 MPL[®] adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the combination of CpG and QS21 is disclosed in WO 00/09159. Preferably the formulation additionally 30 comprises an oil in water emulsion and tocopherol.

Additional illustrative adjuvants for use in the pharmaceutical compositions of the invention include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.*, SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), Detox (Enhazyne[®]) (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties, and polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

10 Other preferred adjuvants include adjuvant molecules of the general formula



wherein, n is 1-50, A is a bond or $-\text{C}(\text{O})-$, R is C_{1-50} alkyl or Phenyl C_{1-50} alkyl.

One embodiment of the present invention consists of a vaccine
15 formulation comprising a polyoxyethylene ether of general formula (I), wherein n is between 1 and 50, preferably 4-24, most preferably 9; the R component is C_{1-50} , preferably C_{4-20} alkyl and most preferably C_{12} alkyl, and A is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene
20 ethers are selected from the following group: polyoxyethylene-9-lauryl ether, polyoxyethylene-9-stearyl ether, polyoxyethylene-8-stearyl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl ether are described in the Merck index (12th edition: entry 7717). These adjuvant molecules are described in WO
25 99/52549.

The polyoxyethylene ether according to the general formula (I) above may, if desired, be combined with another adjuvant. For example, a preferred adjuvant combination is preferably with CpG as described in the pending UK patent application GB 9820956.2.

30 According to another embodiment of this invention, an immunogenic composition described herein is delivered to a host via antigen presenting cells (APCs),

- such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be 5 immunologically compatible with the receiver (*i.e.*, matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic 10 cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature* 392:245-251, 1998) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (see Timmerman and Levy, *Ann. Rev. Med.* 50:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, 15 with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up, process and present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As 20 an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (see Zitvogel et al., *Nature Med.* 4:594-600, 1998).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph 25 nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNF α to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into 30 dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNF α ,

CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fc γ receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (e.g., CD54 and CD11) and costimulatory molecules (e.g., CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with a polynucleotide of the invention (or portion or other variant thereof) such that the encoded polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a pharmaceutical composition comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., *Immunology and cell Biology* 75:456-460, 1997. Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with the tumor polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (e.g., vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (e.g., a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier

will typically vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example, topical, oral, nasal, mucosal, intravenous, intracranial, intraperitoneal, subcutaneous and intramuscular administration.

- 5 Carriers for use within such pharmaceutical compositions are biocompatible, and may also be biodegradable. In certain embodiments, the formulation preferably provides a relatively constant level of active component release. In other embodiments, however, a more rapid rate of release immediately upon administration may be desired. The formulation of such compositions is well within the
10 level of ordinary skill in the art using known techniques. Illustrative carriers useful in this regard include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other illustrative delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (*e.g.*, a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer
15 comprising an amphiphilic compound, such as a phospholipid (*see e.g.*, U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.
- 20 In another illustrative embodiment, biodegradable microspheres (*e.g.*, polylactate polyglycolate) are employed as carriers for the compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763; 5,814,344, 5,407,609 and 5,942,252. Modified hepatitis B core protein carrier systems.
25 such as described in WO/99 40934, and references cited therein, will also be useful for many applications. Another illustrative carrier/delivery system employs a carrier comprising particulate-protein complexes, such as those described in U.S. Patent No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte responses in a host.
- 30 In another illustrative embodiment, calcium phosphate core particles are employed as carriers, vaccine adjuvants, or as controlled release matrices for the

compositions of this invention. Exemplary calcium phosphate particles are disclosed, for example, in published patent application No. WO/0046147.

The pharmaceutical compositions of the invention will often further comprise one or more buffers (*e.g.*, neutral buffered saline or phosphate buffered saline), carbohydrates (*e.g.*, glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (*e.g.*, aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, suspending agents, thickening agents and/or preservatives. Alternatively, 10 compositions of the present invention may be formulated as a lyophilizate.

The pharmaceutical compositions described herein may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers are typically sealed in such a way to preserve the sterility and stability of the formulation until use. In general, formulations may be stored as suspensions, solutions 15 or emulsions in oily or aqueous vehicles. Alternatively, a pharmaceutical composition may be stored in a freeze-dried condition requiring only the addition of a sterile liquid carrier immediately prior to use.

The development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including 20 *e.g.*, oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation, is well known in the art, some of which are briefly discussed below for general purposes of illustration.

In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions 25 may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, 30 suspensions, syrups, wafers, and the like (see, for example, Mathiowitz *et al.*, *Nature* 1997 Mar 27;386(6623):410-4; Hwang *et al.*, *Crit Rev Ther Drug Carrier Syst*

1998;15(3):243-84; U. S. Patent 5,641,515; U. S. Patent 5,580,579 and U. S. Patent 5,792,451). Tablets, troches, pills, capsules and the like may also contain any of a variety of additional components, for example, a binder, such as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, 5 such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as 10 coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

15 Typically, these formulations will contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be varied and may conveniently be between about 1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared is such 20 a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

25 For oral administration the compositions of the present invention may alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. Alternatively, the active ingredient may be incorporated into an oral solution such as one containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a 30 dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants.

Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or otherwise dissolved in the mouth.

In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, intravenously, intramuscularly, or even 5 intraperitoneally. Such approaches are well known to the skilled artisan, some of which are further described, for example, in U. S. Patent 5,543,158; U. S. Patent 5,641,515 and U. S. Patent 5,399,363. In certain embodiments, solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be 10 prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations generally will contain a preservative to prevent the growth of microorganisms.

Illustrative pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous 15 preparation of sterile injectable solutions or dispersions (for example, see U. S. Patent 5,466,468). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium 20 containing, for example, water, ethanol, polyol (e.g., glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and/or by the use of surfactants. The prevention of the action of microorganisms can be 25 facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in 30 the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

In one embodiment, for parenteral administration in an aqueous solution, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal 5 administration. In this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-10 1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. Moreover, for human administration, preparations will of course preferably meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

In another embodiment of the invention, the compositions disclosed 15 herein may be formulated in a neutral or salt form. Illustrative pharmaceutically-acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be 20 derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective.

25 The carriers can further comprise any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active 30 ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions. The phrase

"pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human.

In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles.

- 5 Methods for delivering genes, nucleic acids, and peptide compositions directly to the lungs *via* nasal aerosol sprays has been described, *e.g.*, in U. S. Patent 5,756,353 and U. S. Patent 5,804,212. Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga *et al.*, J Controlled Release 1998 Mar 2;52(1-2):81-7) and lysophosphatidyl-glycerol compounds (U. S. Patent 5,725,871) are also well-known in
10 the pharmaceutical arts. Likewise, illustrative transmucosal drug delivery in the form of a polytetrafluoroethylene support matrix is described in U. S. Patent 5,780,045.

- In certain embodiments, liposomes, nanocapsules, microparticles, lipid particles, vesicles, and the like, are used for the introduction of the compositions of the present invention into suitable host cells/organisms. In particular, the compositions of
15 the present invention may be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, or a nanoparticle or the like. Alternatively, compositions of the present invention can be bound, either covalently or non-covalently, to the surface of such carrier vehicles.

- The formation and use of liposome and liposome-like preparations as
20 potential drug carriers is generally known to those of skill in the art (see for example, Lasic, Trends Biotechnol 1998 Jul;16(7):307-21; Takakura, Nippon Rinsho 1998 Mar;56(3):691-5; Chandran *et al.*, Indian J Exp Biol. 1997 Aug;35(8):801-9; Margalit, Crit Rev Ther Drug Carrier Syst. 1995;12(2-3):233-61; U.S. Patent 5,567,434; U.S. Patent 5,552,157; U.S. Patent 5,565,213; U.S. Patent 5,738,868 and U.S. Patent
25 5,795,587, each specifically incorporated herein by reference in its entirety).

- Liposomes have been used successfully with a number of cell types that are normally difficult to transfet by other procedures, including T cell suspensions, primary hepatocyte cultures and PC 12 cells (Renneisen *et al.*, J Biol Chem. 1990 Sep 25;265(27):16337-42; Muller *et al.*, DNA Cell Biol. 1990 Apr;9(3):221-9). In addition,
30 liposomes are free of the DNA length constraints that are typical of viral-based delivery systems. Liposomes have been used effectively to introduce genes, various drugs,

radiotherapeutic agents, enzymes, viruses, transcription factors, allosteric effectors and the like, into a variety of cultured cell lines and animals. Furthermore, the use of liposomes does not appear to be associated with autoimmune responses or unacceptable toxicity after systemic delivery.

5 In certain embodiments, liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs).

Alternatively, in other embodiments, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the 10 present invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (see, for example, Quintanar-Guerrero *et al.*, Drug Dev Ind Pharm. 1998 Dec;24(12):1113-28). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) may be designed using polymers able to be degraded *in vivo*. Such particles can be made as described, for 15 example, by Couvreur *et al.*, Crit Rev Ther Drug Carrier Syst. 1988;5(1):1-20; zur Muhlen *et al.*, Eur J Pharm Biopharm. 1998 Mar;45(2):149-55; Zambaux *et al.* J Controlled Release. 1998 Jan 2;50(1-3):31-40; and U. S. Patent 5,145,684.

Cancer Therapeutic Methods

Immunologic approaches to cancer therapy are based on the recognition 20 that cancer cells can often evade the body's defenses against aberrant or foreign cells and molecules, and that these defenses might be therapeutically stimulated to regain the lost ground, e.g. pgs. 623-648 in Klein, Immunology (Wiley-Interscience, New York, 1982). Numerous recent observations that various immune effectors can directly or indirectly inhibit growth of tumors has led to renewed interest in this approach to cancer 25 therapy, e.g. Jager, et al., Oncology 2001;60(1):1-7; Renner, et al., Ann Hematol 2000 Dec;79(12):651-9.

Four-basic cell types whose function has been associated with antitumor cell immunity and the elimination of tumor cells from the body are: i) B-lymphocytes which secrete immunoglobulins into the blood plasma for identifying and labeling the 30 nonself invader cells; ii) monocytes which secrete the complement proteins that are

- responsible for lysing and processing the immunoglobulin-coated target invader cells; iii) natural killer lymphocytes having two mechanisms for the destruction of tumor cells, antibody-dependent cellular cytotoxicity and natural killing; and iv) T-lymphocytes possessing antigen-specific receptors and having the capacity to recognize
5 a tumor cell carrying complementary marker molecules (Schreiber, H., 1989, in Fundamental Immunology (ed). W. E. Paul, pp. 923-955).

Cancer immunotherapy generally focuses on inducing humoral immune responses, cellular immune responses, or both. Moreover, it is well established that induction of CD4⁺ T helper cells is necessary in order to secondarily induce either
10 antibodies or cytotoxic CD8⁺ T cells. Polypeptide antigens that are selective or ideally specific for cancer cells, particularly lung cancer cells, offer a powerful approach for inducing immune responses against lung cancer, and are an important aspect of the present invention.

Therefore, in further aspects of the present invention, the pharmaceutical
15 compositions described herein may be used to stimulate an immune response against cancer, particularly for the immunotherapy of lung cancer. Within such methods, the pharmaceutical compositions described herein are administered to a patient, typically a warm-blooded animal, preferably a human. A patient may or may not be afflicted with cancer. Pharmaceutical compositions and vaccines may be administered either prior to
20 or following surgical removal of primary tumors and/or treatment such as administration of radiotherapy or conventional chemotherapeutic drugs. As discussed above, administration of the pharmaceutical compositions may be by any suitable method, including administration by intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, intradermal, anal, vaginal, topical and oral routes.

25 Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against tumors with the administration of immune response-modifying agents (such as polypeptides and polynucleotides as provided herein).

30 Within other embodiments, immunotherapy may be passive immunotherapy, in which treatment involves the delivery of agents with established

tumor-immune reactivity (such as effector cells or antibodies) that can directly or indirectly mediate antitumor effects and does not necessarily depend on an intact host immune system. Examples of effector cells include T cells as discussed above, T lymphocytes (such as CD8⁺ cytotoxic T lymphocytes and CD4⁺ T-helper tumor-

5 infiltrating lymphocytes), killer cells (such as Natural Killer cells and lymphokine-activated killer cells), B cells and antigen-presenting cells (such as dendritic cells and macrophages) expressing a polypeptide provided herein. T cell receptors and antibody receptors specific for the polypeptides recited herein may be cloned, expressed and transferred into other vectors or effector cells for adoptive immunotherapy. The

10 polypeptides provided herein may also be used to generate antibodies or anti-idiotypic antibodies (as described above and in U.S. Patent No. 4,918,164) for passive immunotherapy.

Monoclonal antibodies may be labeled with any of a variety of labels for desired selective usages in detection, diagnostic assays or therapeutic applications (as

15 described in U.S. Patent Nos. 6,090,365; 6,015,542; 5,843,398; 5,595,721; and 4,708,930, hereby incorporated by reference in their entirety as if each was incorporated individually). In each case, the binding of the labelled monoclonal antibody to the determinant site of the antigen will signal detection or delivery of a particular therapeutic agent to the antigenic determinant on the non-normal cell. A further object

20 of this invention is to provide the specific monoclonal antibody suitably labelled for achieving such desired selective usages thereof.

Effector cells may generally be obtained in sufficient quantities for adoptive immunotherapy by growth *in vitro*, as described herein. Culture conditions for expanding single antigen-specific effector cells to several billion in number with

25 retention of antigen recognition *in vivo* are well known in the art. Such *in vitro* culture conditions typically use intermittent stimulation with antigen, often in the presence of cytokines (such as IL-2) and non-dividing feeder cells. As noted above, immunoreactive polypeptides as provided herein may be used to rapidly expand antigen-specific T cell cultures in order to generate a sufficient number of cells for

30 immunotherapy. In particular, antigen-presenting cells, such as dendritic, macrophage, monocyte, fibroblast and/or B cells, may be pulsed with immunoreactive polypeptides

or transfected with one or more polynucleotides using standard techniques well known in the art. For example, antigen-presenting cells can be transfected with a polynucleotide having a promoter appropriate for increasing expression in a recombinant virus or other expression system. Cultured effector cells for use in therapy 5 must be able to grow and distribute widely, and to survive long term *in vivo*. Studies have shown that cultured effector cells can be induced to grow *in vivo* and to survive long term in substantial numbers by repeated stimulation with antigen supplemented with IL-2 (see, for example, Cheever et al., *Immunological Reviews* 157:177, 1997).

Alternatively, a vector expressing a polypeptide recited herein may be 10 introduced into antigen presenting cells taken from a patient and clonally propagated *ex vivo* for transplant back into the same patient. Transfected cells may be reintroduced into the patient using any means known in the art, preferably in sterile form by intravenous, intracavitory, intraperitoneal or intratumor administration.

Routes and frequency of administration of the therapeutic compositions 15 described herein, as well as dosage, will vary from individual to individual, and may be readily established using standard techniques. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g.*, intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Preferably, between 1 and 10 doses may be administered over a 52 week period. 20 Preferably, 6 doses are administered, at intervals of 1 month, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of a compound that, when administered as described above, is capable of promoting an anti-tumor immune response, and is at least 10-50% above the basal (*i.e.*, untreated) level. Such response 25 can be monitored by measuring the anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient's tumor cells *in vitro*. Such vaccines should also be capable of causing an immune response that leads to an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non- 30 vaccinated patients. In general, for pharmaceutical compositions and vaccines comprising one or more polypeptides, the amount of each polypeptide present in a dose

ranges from about 25 µg to 5 mg per kg of host. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

In general, an appropriate dosage and treatment regimen provides the active compound(s) in an amount sufficient to provide therapeutic and/or prophylactic benefit. Such a response can be monitored by establishing an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial, or longer disease-free survival) in treated patients as compared to non-treated patients. Increases in preexisting immune responses to a tumor protein generally correlate with an improved clinical outcome. Such immune responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using samples obtained from a patient before and after treatment.

Cancer Detection and Diagnostic Compositions, Methods and Kits

In general, a cancer may be detected in a patient based on the presence of one or more lung tumor proteins and/or polynucleotides encoding such proteins in a biological sample (for example, blood, sera, sputum urine and/or tumor biopsies) obtained from the patient. In other words, such proteins may be used as markers to indicate the presence or absence of a cancer such as lung cancer. In addition, such proteins may be useful for the detection of other cancers. The binding agents provided herein generally permit detection of the level of antigen that binds to the agent in the biological sample.

Polynucleotide primers and probes may be used to detect the level of mRNA encoding a tumor protein, which is also indicative of the presence or absence of a cancer. In general, a tumor sequence should be present at a level that is at least two-fold, preferably three-fold, and more preferably five-fold or higher in tumor tissue than in normal tissue of the same type from which the tumor arose. Expression levels of a particular tumor sequence in tissue types different from that in which the tumor arose are irrelevant in certain diagnostic embodiments since the presence of tumor cells can be confirmed by observation of predetermined differential expression levels, *e.g.*, 2-fold, 5-fold, etc, in tumor tissue to expression levels in normal tissue of the same type.

Other differential expression patterns can be utilized advantageously for diagnostic purposes. For example, in one aspect of the invention, overexpression of a tumor sequence in tumor tissue and normal tissue of the same type, but not in other normal tissue types, e.g. PBMCs, can be exploited diagnostically. In this case, the
5 presence of metastatic tumor cells, for example in a sample taken from the circulation or some other tissue site different from that in which the tumor arose, can be identified and/or confirmed by detecting expression of the tumor sequence in the sample, for example using RT-PCR analysis. In many instances, it will be desired to enrich for tumor cells in the sample of interest, e.g., PBMCs, using cell capture or other like
10 techniques.

There are a variety of assay formats known to those of ordinary skill in the art for using a binding agent to detect polypeptide markers in a sample. See, e.g., Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, the presence or absence of a cancer in a patient may be determined by
15 (a) contacting a biological sample obtained from a patient with a binding agent; (b) detecting in the sample a level of polypeptide that binds to the binding agent; and (c) comparing the level of polypeptide with a predetermined cut-off value.

In a preferred embodiment, the assay involves the use of binding agent immobilized on a solid support to bind to and remove the polypeptide from the
20 remainder of the sample. The bound polypeptide may then be detected using a detection reagent that contains a reporter group and specifically binds to the binding agent/polypeptide complex. Such detection reagents may comprise, for example, a binding agent that specifically binds to the polypeptide or an antibody or other agent that specifically binds to the binding agent, such as an anti-immunoglobulin, protein G,
25 protein A or a lectin. Alternatively, a competitive assay may be utilized, in which a polypeptide is labeled with a reporter group and allowed to bind to the immobilized binding agent after incubation of the binding agent with the sample. The extent to which components of the sample inhibit the binding of the labeled polypeptide to the binding agent is indicative of the reactivity of the sample with the immobilized binding
30 agent. Suitable polypeptides for use within such assays include full length lung tumor

proteins and polypeptide portions thereof to which the binding agent binds, as described above.

- The solid support may be any material known to those of ordinary skill in the art to which the tumor protein may be attached. For example, the solid support
- 5 may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681. The binding agent may be immobilized on the solid support
- 10 using a variety of techniques known to those of skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "immobilization" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the agent and functional groups on the support or may be a linkage by way of a cross-linking agent).
- 15 Immobilization by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the binding agent, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and about 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or
- 20 polyvinylchloride) with an amount of binding agent ranging from about 10 ng to about 10 µg, and preferably about 100 ng to about 1 µg, is sufficient to immobilize an adequate amount of binding agent.

Covalent attachment of binding agent to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with

25 both the support and a functional group, such as a hydroxyl or amino group, on the binding agent. For example, the binding agent may be covalently attached to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the binding partner (see, e.g., Pierce Immunotechnology Catalog and Handbook, 1991, at

30 A12-A13).

- In certain embodiments, the assay is a two-antibody sandwich assay. This assay may be performed by first contacting an antibody that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that polypeptides within the sample are allowed to bind to the immobilized antibody.
- 5 Unbound sample is then removed from the immobilized polypeptide-antibody complexes and a detection reagent (preferably a second antibody capable of binding to a different site on the polypeptide) containing a reporter group is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific reporter group.
- 10 More specifically, once the antibody is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20TM (Sigma Chemical Co., St. Louis, MO). The immobilized antibody is then incubated with the sample, and polypeptide is allowed to
- 15 bind to the antibody. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is a period of time that is sufficient to detect the presence of polypeptide within a sample obtained from an individual with lung least about 95% of that achieved at equilibrium between bound and unbound polypeptide. Those of
- 20 ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.
- Unbound sample may then be removed by washing the solid support
- 25 with an appropriate buffer, such as PBS containing 0.1% Tween 20TM. The second antibody, which contains a reporter group, may then be added to the solid support. Preferred reporter groups include those groups recited above.
- The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound polypeptide.
- 30 An appropriate amount of time may generally be determined by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed

and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups
5 and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

10 To determine the presence or absence of a cancer, such as lung cancer, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value for the detection of a cancer is the average mean signal obtained when the immobilized antibody is incubated with samples from
15 patients without the cancer. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value is considered positive for the cancer. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985,
20 p. 106-7. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a
25 signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for a cancer.

30 In a related embodiment, the assay is performed in a flow-through or strip test format, wherein the binding agent is immobilized on a membrane, such as

nitrocellulose. In the flow-through test, polypeptides within the sample bind to the immobilized binding agent as the sample passes through the membrane. A second, labeled binding agent then binds to the binding agent-polypeptide complex as a solution containing the second binding agent flows through the membrane. The detection of 5 bound second binding agent may then be performed as described above. In the strip test format, one end of the membrane to which binding agent is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing second binding agent and to the area of immobilized binding agent. Concentration of second binding agent at the area of immobilized antibody indicates the 10 presence of a cancer. Typically, the concentration of second binding agent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of binding agent immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of polypeptide that would be sufficient to generate a 15 positive signal in the two-antibody sandwich assay, in the format discussed above. Preferred binding agents for use in such assays are antibodies and antigen-binding fragments thereof. Preferably, the amount of antibody immobilized on the membrane ranges from about 25 ng to about 1 μ g, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount of biological 20 sample.

Of course, numerous other assay protocols exist that are suitable for use with the tumor proteins or binding agents of the present invention. The above descriptions are intended to be exemplary only. For example, it will be apparent to those of ordinary skill in the art that the above protocols may be readily modified to use 25 tumor polypeptides to detect antibodies that bind to such polypeptides in a biological sample. The detection of such tumor protein specific antibodies may correlate with the presence of a cancer.

A cancer may also, or alternatively, be detected based on the presence of T cells that specifically react with a tumor protein in a biological sample. Within 30 certain methods, a biological sample comprising CD4 $^{+}$ and/or CD8 $^{+}$ T cells isolated from a patient is incubated with a tumor polypeptide, a polynucleotide encoding such a

- polypeptide and/or an APC that expresses at least an immunogenic portion of such a polypeptide, and the presence or absence of specific activation of the T cells is detected. Suitable biological samples include, but are not limited to, isolated T cells. For example, T cells may be isolated from a patient by routine techniques (such as by
- 5 Ficoll/Hypaque density gradient centrifugation of peripheral blood lymphocytes). T cells may be incubated *in vitro* for 2-9 days (typically 4 days) at 37°C with polypeptide (e.g., 5 - 25 µg/ml). It may be desirable to incubate another aliquot of a T cell sample in the absence of tumor polypeptide to serve as a control. For CD4⁺ T cells, activation is preferably detected by evaluating proliferation of the T cells. For CD8⁺ T cells,
- 10 activation is preferably detected by evaluating cytolytic activity. A level of proliferation that is at least two fold greater and/or a level of cytolytic activity that is at least 20% greater than in disease-free patients indicates the presence of a cancer in the patient.

As noted above, a cancer may also, or alternatively, be detected based on the level of mRNA encoding a tumor protein in a biological sample. For example, at

15 least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a tumor cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (*i.e.*, hybridizes to) a polynucleotide encoding the tumor protein. The amplified cDNA is then separated and detected using techniques well known in the art, such as gel

20 electrophoresis.

Similarly, oligonucleotide probes that specifically hybridize to a polynucleotide encoding a tumor protein may be used in a hybridization assay to detect the presence of polynucleotide encoding the tumor protein in a biological sample.

To permit hybridization under assay conditions, oligonucleotide primers

25 and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a tumor protein of the invention that is at least 10 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers and/or probes hybridize to a polynucleotide encoding a

30 polypeptide described herein under moderately stringent conditions, as defined above. Oligonucleotide primers and/or probes which may be usefully employed in the

- diagnostic methods described herein preferably are at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having a sequence as disclosed herein. Techniques for both PCR based assays and 5 hybridization assays are well known in the art (see, for example, Mullis et al., *Cold Spring Harbor Symp. Quant. Biol.*, 51:263, 1987; Erlich ed., *PCR Technology*, Stockton Press, NY, 1989).

One preferred assay employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological 10 sample, such as biopsy tissue, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a test patient and from an individual who is not afflicted with a cancer. The amplification reaction may be 15 performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

In another aspect of the present invention, cell capture technologies may 20 be used in conjunction, with, for example, real-time PCR to provide a more sensitive tool for detection of metastatic cells expressing lung tumor antigens. Detection of lung cancer cells in biological samples, e.g., bone marrow samples, peripheral blood, and small needle aspiration samples is desirable for diagnosis and prognosis in lung cancer patients.

25 Immunomagnetic beads coated with specific monoclonal antibodies to surface cell markers, or tetrameric antibody complexes, may be used to first enrich or positively select cancer cells in a sample. Various commercially available kits may be used, including Dynabeads® Epithelial Enrich (Dynal Biotech, Oslo, Norway), StemSep™ (StemCell Technologies, Inc., Vancouver, BC), and RosetteSep (StemCell 30 Technologies). A skilled artisan will recognize that other methodologies and kits may also be used to enrich or positively select desired cell populations. Dynabeads®

Epithelial Enrich contains magnetic beads coated with mAbs specific for two glycoprotein membrane antigens expressed on normal and neoplastic epithelial tissues. The coated beads may be added to a sample and the sample then applied to a magnet, thereby capturing the cells bound to the beads. The unwanted cells are washed away
5 and the magnetically isolated cells eluted from the beads and used in further analyses.

RosetteSep can be used to enrich cells directly from a blood sample and consists of a cocktail of tetrameric antibodies that targets a variety of unwanted cells and crosslinks them to glycophorin A on red blood cells (RBC) present in the sample, forming rosettes. When centrifuged over Ficoll, targeted cells pellet along with the free
10 RBC. The combination of antibodies in the depletion cocktail determines which cells will be removed and consequently which cells will be recovered. Antibodies that are available include, but are not limited to: CD2, CD3, CD4, CD5, CD8, CD10, CD11b, CD14, CD15, CD16, CD19, CD20, CD24, CD25, CD29, CD33, CD34, CD36, CD38, CD41, CD45, CD45RA, CD45RO, CD56, CD66B, CD66e, HLA-DR, IgE, and TCR $\alpha\beta$.
15

Additionally, it is contemplated in the present invention that mAbs specific for lung tumor antigens can be generated and used in a similar manner. For example, mAbs that bind to tumor-specific cell surface antigens may be conjugated to magnetic beads, or formulated in a tetrameric antibody complex, and used to enrich or positively select metastatic lung tumor cells from a sample. Once a sample is enriched
20 or positively selected, cells may be lysed and RNA isolated. RNA may then be subjected to RT-PCR analysis using lung tumor-specific primers in a real-time PCR assay as described herein. One skilled in the art will recognize that enriched or selected populations of cells may be analyzed by other methods (*e.g.* *in situ* hybridization or flow cytometry).

In another embodiment, the compositions described herein may be used as markers for the progression of cancer. In this embodiment, assays as described above for the diagnosis of a cancer may be performed over time, and the change in the level of reactive polypeptide(s) or polynucleotide(s) evaluated. For example, the assays may be performed every 24-72 hours for a period of 6 months to 1 year, and thereafter
30 performed as needed. In general, a cancer is progressing in those patients in whom the level of polypeptide or polynucleotide detected increases over time. In contrast, the

cancer is not progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

Certain *in vivo* diagnostic assays may be performed directly on a tumor. One such assay involves contacting tumor cells with a binding agent. The bound 5 binding agent may then be detected directly or indirectly via a reporter group. Such binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

As noted above, to improve sensitivity, multiple tumor protein markers may be assayed within a given sample. It will be apparent that binding agents specific 10 for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based on routine experiments to determine combinations that results in optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

15 The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a tumor protein. 20 Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

25 Alternatively, a kit may be designed to detect the level of mRNA encoding a tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a tumor protein. Such an oligonucleotide may be used, for example, within a PCR or hybridization assay. Additional components that may be 30 present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a tumor protein.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

5

EXAMPLE 1

IDENTIFICATION OF LUNG TUMOR PROTEIN cDNAs

Lung-specific genes were identified by electronic subtraction. The method used was similar to that described by Vasmatzis et al., *Proc. Natl. Acad. Sci. USA* 95:300-304, 1998, but there were several key differences. Sequences of EST clones (1,453,679) were downloaded from the GenBank public human EST database. Human cDNA libraries were downloaded to create a database of these cDNA libraries and the EST sequences derived from them. The cDNA libraries were grouped into three groups: Plus, Minus and Other/Neutral. The Plus group included 30 libraries constructed from lung tumor and fetal lung tissues (and therefore including those containing lung tumor-specific ESTs); the Minus group consisted of 206 libraries derived from all adult normal tissues; the Other/Neutral group contained libraries from tissues where expression is considered irrelevant (*e.g.*, non-lung-fetal tissue, non-lung tumors, cell lines other than lung tumor cell lines). A total of 93,526 ESTs were derived from the 30 lung tumor and fetal lung libraries. These ESTs were preprocessed to remove common sequence repeats and cloning adapters, resulting in a final Plus group of 90,365 (a decrease of 3%).

Each Plus group (lung tumor or fetal lung) EST sequence was used as a query “seed” sequence in a BLASTN (version 2.0.9; May 7, 1999) search against the total human EST database. Standard measures of similarity are insufficient in this sort of analysis, as EST relationships often include short stretches and poor sequence data. Criteria employed in this study required a matching segment to be at least 75 nucleotides in length, and the density of exact matches within this segment to be at least 80%. This was considered conservative criteria designed to avoid short spurious matches while allowing for polymorphisms and errors in sequencing. Each BLAST search generated a cluster of related sequences based on direct overlap with the query

"seed" sequence. A second level of clustering was performed to merge closely related clusters and to eliminate redundancy resulting from the fact that similar clusters are generated if the clusters contain more than one seed (*i.e.*, sequences from the Plus EST group). The resulting "super clusters" were discarded if they grew in size to 200 or 5 more ESTs, since these probably represented repetitive elements that were not removed by the initial preprocessing of the seeds, or highly expressed genes such as those for ribosomal proteins. Superclusters were merged if they shared at least one third of their sequences.

The BLAST searches gave rise to a total of 49,154 clusters. In the first 10 super clustering stage, 18,665 clusters grew beyond the limit of 200 clones. The remainder was reduced to a total of 30,489 super clusters. This number was reduced to 29,501 after adjacent clusters were merged. Resulting super clusters were analyzed to determine the tissue source of each EST clone contained within it and this expression profile was used to classify the superclusters into four groups: Type 1 – this 15 supercluster contains EST clones found in the Plus group only, with no expression in the Minus or Other/Neutral group libraries; Type 2 – EST clones in the supercluster are found in the Plus and Other/Neutral group libraries, with no expression in the Minus group; Type 3 – super cluster EST clones found in all groups, but the number of ESTs in the Plus group is higher than in either of the Minus or Other/Neutral groups; Type 4 – 20 super cluster EST clones found in all groups, but the number in the Plus group is higher than in the Minus group with expression in the Other/Neutral group non relevant. Sequences derived from the Plus library group that were placed in Types 1, 2 and 3 superclusters resulted in 20,487 polynucleotide sequences. The electronic subtraction procedures identified these sequences as having significant differential expression in 25 lung tissue.

EXAMPLE 2

ANALYSIS OF cDNA EXPRESSION USING MICROARRAY TECHNOLOGY

2208 of the clones identified from the lung electronic subtraction procedure were evaluated for overexpression in specific tumor tissues by microarray 30 analysis. Using this approach, cDNA sequences are PCR amplified and their mRNA expression profiles in tumor and normal tissues are examined using cDNA microarray

technology essentially as described (Shena, M. *et al.*, 1995 *Science* 270:467-70). In brief, the 2208 clones were arrayed onto glass slides as multiple replicas, with each location corresponding to a unique cDNA clone (as many as 5500 clones can be arrayed on a single slide or chip). Each chip was hybridized with a pair of cDNA probes that
5 were fluorescence-labeled with Cy3 and Cy5, respectively. Typically, 1 μ g of polyA⁺ RNA was used to generate each cDNA probe. Since one cDNA probe is generated from tumor tissue RNA and the other is generated from normal tissue RNA, sequences that are differentially overexpressed in tumor tissue will generate a stronger signal from the tumor specific probe than the normal tissue probe, thus allowing the identification of
10 those sequences that exhibit elevated expression in tumor versus normal tissue.

After hybridization, the chips were scanned and the fluorescence intensity recorded for both Cy3 and Cy5 channels. There were multiple built-in quality control steps. First, the probe quality was monitored using a panel of 18 ubiquitously expressed genes. Secondly, the control plate also had yeast DNA fragments of which
15 complementary RNA was spiked into the probe synthesis for measuring the quality of the probe and the sensitivity of the analysis. Currently, the technology offers a sensitivity of 1 in 100,000 copies of mRNA. Finally, the reproducibility of this technology was ensured by including duplicated control cDNA elements at different locations. Further validation of the process was indicated in that several differentially
20 expressed genes were identified multiple times in the study, and the expression profiles for these genes are very comparable. The clones were arrayed on Lung Chip 6.

Of those analyzed by microarray, 781 sequences met the criteria of having at least 2-fold overexpression in lung tumor tissue compared to normal tissues. Of these 781 clones, 459 were found to meet the additional criteria of having a mean
25 normal tissue expression value less than or equal to 0.2. These 459 clones were then analyzed visually and certain clones with favorable expression profiles (e.g., high expression in tumors with little or no expression in normal tissues) were sequenced and searched against public sequence databases to facilitate identification of extended sequence for the clones.

30 SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 32 and
34 represent a subset of those 459 clones that met the above criteria of being at least 2-

fold overexpressed in tumor versus normal tissues and having a mean normal tissue expression of less than or equal to 0.2. Additional information about these sequences is provided in Table 2 below.

Table 2

| SEQ ID NO: | SEQ ID NO: from 60/207,485 | Clone Name: | Clone ID # | MICROARRAY ANALYSIS (Lung Chip #) | MICROARRAY RATIO (Lung Tumor:Normal Tissue) |
|------------|----------------------------------|-------------|------------|--------------------------------------|--|
| 9 | 4538 | L1027C | 55571 | 6 | 2.94 |
| 5 | 4978 | L1037C | 58267 | 6 | 2.61 |
| 7 | 1796 | L1038C | 58245 | 6 | 3.5 |
| 3 | 7264 | L1039C | 58269 | 6 | 2.81 |
| 1 | 2337 | L1040C | 55964 | 6 | 5.07 |
| 15 | 1548/4619 | L1041C | 58346 | 6 | 2.33 |
| 25 | 15127 | n/a | 56016 | 6 | > 2 |
| 27 | 3816 | n/a | 55987 | 6 | > 2 |
| 29 | 2046 | n/a | 55956 | 6 | > 2 |
| 31 | 1912 | n/a | 55952 | 6 | > 2 |
| 32 | 2064 | n/a | 55957 | 6 | > 2 |
| 34 | 1502/3852 | n/a | 55559 | 6 | > 2 |
| 11 | 2814 | n/a | 55978 | 6 | > 2 |
| 13 | 3478 | n/a | 55980 | 6 | > 2 |
| 17 | 553 | n/a | 55561 | 6 | > 2 |
| 19 | 3275 | n/a | 55984 | 6 | > 2 |
| 21 | 2809 | n/a | 58261 | 6 | > 2 |
| 23 | 1677 | n/a | 58348 | 6 | > 2 |

Each of the sequences was then used as a query to search the public databases in order to facilitate identification of extended sequences for these clones. Extended sequence information for the above sequences, obtained by searching public sequence databases, is set forth in SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 5 26, 28, 30, 33, and 35, respectively.

EXAMPLE 3

QUANTITATIVE REAL-TIME RT-PCR ANALYSIS

Briefly, quantitation of PCR product relies on the few cycles where the 10 amount of DNA amplifies logarithmically from barely above the background to the plateau. Using continuous fluorescence monitoring, the threshold cycle number where DNA amplifies logarithmically is easily determined in each PCR reaction. There are two fluorescence detecting systems. One is based upon a double-strand DNA specific binding dye SYBR Green I dye. The other uses TaqMan probe containing a Reporter 15 dye at the 5' end (FAM) and a Quencher dye at the 3' end (TAMRA) (Perkin Elmer/Applied Biosystems Division, Foster City, CA). Target-specific PCR amplification results in cleavage and release of the Reporter dye from the Quencher-containing probe by the nuclease activity of AmpliTaq GoldTM (Perkin Elmer/Applied Biosystems Division, Foster City, CA). Thus, fluorescence signal generated from 20 released reporter dye is proportional to the amount of PCR product. Both detection methods have been found to generate comparable results. To compare the relative level of gene expression in multiple tissue samples, a panel of cDNAs is constructed using RNA from tissues and/or cell lines, and Real-Time PCR is performed using gene specific primers to quantify the copy number in each cDNA sample. Each cDNA 25 sample is generally performed in duplicate and each reaction repeated in duplicated plates. The final Real-time PCR result is typically reported as an average of copy number of a gene of interest normalized against internal actin number in each cDNA sample. Real-time PCR reactions may be performed on a GeneAmp 5700 Detector using SYBR Green I dye or an ABI PRISM 7700 Detector using the TaqMan probe 30 (Perkin Elmer/Applied Biosystems Division, Foster City, CA).

Using this approach, Real Time PCR® profiles were generated for L1027, L1037, L1038, L1039, L1040 and L1041, and are provided in Table 3.

Table 3

| SEQ ID NO: | CLONE NAME | REAL TIME PROFILE |
|------------|------------|--|
| 9 | L1027C | Real Time PCR shows over-expression in small cell lung carcinoma as well as in bone marrow. Expression is also observed for multiple normal tissue. |
| 5 | L1037C | Real Time PCR shows over-expression in small cell lung carcinoma as well as in bone marrow and lymph node. Expression is also observed for multiple normal tissue. |
| 7 | L1038C | Real Time PCR shows over-expression in small cell lung carcinoma as well as in brain, pituitary gland and adrenal gland. Expression is also observed for multiple normal tissue. |
| 3 | L1039C | Real Time PCR shows over-expression in small cell lung carcinoma as well as in lymph node. Expression is also observed for multiple normal tissue. |
| 1 | L1040C | Real Time PCR shows over-expression in small cell lung carcinoma as well as in brain, pituitary gland and adrenal gland. Expression is also observed for multiple normal tissue. |
| 15 | L0141C | Real Time PCR shows over-expression in small cell lung carcinoma as well as in adrenal gland, bone marrow and thymus. Expression is also observed for multiple normal tissue. |

EXAMPLE 4

CLONING OF FULL-LENGTH cDNA SEQUENCES AND ORF FOR L1027C

cDNA sequences encoding the full-length sequence for L1027C were isolated by screening a small cell primary tumor full length cloning library with a 10 radioactively labeled probe of the original isolate sequence (SEQ ID NO:9). In order to determine the transcript size of the gene, a multiple tissue Northern blot was probed with the radioactively labelled original isolate sequence, SEQ ID NO:9. The Northern blot included 1µg of small cell primary tumor polyA+ RNA. Visual analysis of the

exposed film revealed a single transcript of approximately 2.5 kb. Approximately 500,000 clones from the full-length cloning library were screened and four clones were obtained from this library. The inserts were sequenced and yielded DNA nucleotide molecules of about 2.32 and 2.37 kb. These sequences are provided in SEQ ID NO:93
5 and 94, respectively. Both of these sequences contain the same single OFR of 450 bp (SEQ ID NO:95), and encode a deduced amino acid sequence of 150 amino acid residues (SEQ ID NO:96). These sequences were searched against the Genbank nonredundant and GeneSeq DNA databases and showed no hits.

10

EXAMPLE 5

ANALYSIS OF cDNA EXPRESSION USING MICROARRAY TECHNOLOGY

An additional 5054 of the resulting clones obtained from the lung electronic subtraction of Example 1 were probed by microarray chip technology to
15 further characterize the expression of these clones. The microarray analysis was carried out as provided in Example 2. The clones were arrayed on Lung Chip 7. CorixArray analysis was performed on the microarray results to compare expression in lung tumors and in normal tissues. Clones were selected based on two criteria: 2-fold overexpression in lung tumors when compared to non-lung tissue and a mean
20 expression level of less than 0.2 in these same non-lung tissues. Of those analyzed, 2372 clones met the criteria.

Microarray analysis for five of these clones is presented in Table 4:

Table 4

| SEQ ID NO: | SEQ ID NO: from 60/207,485 | Clone Name: | Clone ID # | MICROARRAY ANALYSIS (Lung Chip #) | MICROARRAY RATIO (Lung Tumor:Normal Tissue) |
|------------|----------------------------------|----------------|------------|---|--|
| 42 | 18618 | L1053C | 63575 | 7 | 13.5 |
| 43 | 14788 | L1054C | 63582 | 7 | 5.29 |
| 44 | 7744 | L1055C | 63598 | 7 | 15.25 |
| 45 | 4257 | L1056C | 64963 | 7 | 9.31 |
| 46 | 20087 | L1058C | 64988 | 7 | 5.66 |

5

EXAMPLE 6

QUANTITATIVE REAL-TIME PCR ANALYSIS

- 170 of the 2372 clones of Example 4 were further analyzed by visual analysis based on high expression in tumors and little or no expression in normal tissues. Seven clones were selected for Real-time PCR analysis. The Real-time PCR 10 was carried out as disclosed in Example 3. The Real-time PCR profiles of these seven clones are presented in Table 5. The sequences of these seven clones are provided in SEQ ID NO:42-48, respectively.

Table 5

| SEQ ID NO: | CLONE NAME | CLONE ID # | REAL TIME PROFILE |
|------------|------------|------------|--|
| 42 | L1053C | 63575 | Real Time PCR shows over-expression in small cell lung carcinoma as well as in pituitary. Expression is also observed for multiple normal tissues. |
| 43 | L1054C | 63582 | Real Time PCR shows over-expression in small cell lung carcinoma as well as in pituitary, brain and spinal cord. Expression is also observed for adrenal and pancreas. |
| 44 | L1055C | 63598 | Real Time PCR shows over-expression in small |

| | | | |
|----|--------|-------|---|
| | | | cell lung carcinoma as well as in pituitary and brain. Expression is also observed for multiple normal tissues. |
| 45 | L1056C | 64963 | Real Time PCR shows over-expression in one small cell lung carcinoma sample. No expression is otherwise observed. |
| 46 | L1058C | 64988 | Real Time PCR shows over-expression in small cell lung carcinoma. Low level expression is also observed for adrenal gland, pancreas, and bone marrow. |
| 47 | n/a | 63485 | Real Time PCR shows over-expression in metastatic tumor as well as low level expression in multiple normal tissues. |
| 48 | n/a | 65010 | Real Time PCR shows low expression in one lung sample. No expression is otherwise observed. |

Each of the sequences was then used as a query to search the public databases in order to facilitate identification of extended sequences for these clones. SEQ ID NO:42, 43 and 45 matched to known genes in Genbank, and these results are presented in Table 6. The full-length cDNA sequences of the known genes are disclosed in SEQ ID NO:49, 50 and 52, respectively. The deduced amino acid sequences encoded by SEQ ID NO:49 and 50 are also provided as SEQ ID NO:56 and 57, respectively. SEQ ID NO:44 and 46-48 were found to be novel with respect to known genes, but matched to public EST sequences. The sequences of SEQ ID NO:44 and 46-48 were aligned with the matching EST sequences in order to obtain extended sequence data. These extended sequences are provided in SEQ ID NO:51 and 53-55, respectively.

Table 6

| SEQ ID NO: | CLONE NAME | GENBANK DESCRIPTION |
|------------|------------|-------------------------|
| 42 | L1053C | Insulinoma-associated 1 |
| 43 | L1054C | KIAA0535 |
| 45 | L1056C | Human DAZ mRNA 3' UTR |

EXAMPLE 7

CLONING OF cDNA ENCODING FULL-LENGTH L1058C

The cDNA sequence encoding full-length L1058C was isolated by screening a small cell primary tumor full length cloning library with a radioactively labeled probe of the original isolate sequence (SEQ ID NO:46). In order to determine the transcript size of the gene, a multiple tissue Northern blot was probed with the radioactively labelled original isolate sequence, SEQ ID NO:46. The Northern blot included 1 μ g of small cell primary tumor, carcinoid metastasis and small cell (tumor) cell line polyA+ RNA. Visual analysis of the exposed film revealed a single transcript 5 of approximately 2.5 kb. Approximately 500,000 clones from the full-length cloning library were screened and one clone was obtained from this library. The insert was sequenced and yields a 2165 bp DNA nucleotide molecule. The full-length sequence is provided in SEQ ID NO:58. The full-length sequence is predicted to have two ORFs. A first ORF (SEQ ID NO:59) is predicted to encode a polypeptide having 392 amino 10 acid residues (SEQ ID NO:61), and the second ORF (SEQ ID NO:60) is predicted to encode a polypeptide of 363 amino acid residues (SEQ ID NO:62) but does not show the starting methionine. This 2165 bp DNA was searched against the Genbank nonredundant and GeneSeq DNA databases and showed no hits.

EXAMPLE 8

ANALYSIS OF cDNA EXPRESSION USING MICROARRAY TECHNOLOGY

An additional 3453 of the resulting clones obtained from the lung electronic subtraction of Example 1 were probed by microarray chip technology to further characterize the expression of these clones. The microarray analysis was carried 25 out as provided in Example 2. The clones were arrayed on Lung Chip 8. CorixArray analysis was performed on the microarray results to compare expression in lung tumors and in normal tissues. Clones were selected based on two criteria: 2-fold overexpression in lung tumors when compared to non-lung tissue and a mean expression level of less than 0.2 in these same non-lung tissues. Of those analyzed, 557 30 clones met the criteria.

300 of the 557 clones were visually analyzed for overexpression in tumor versus normal tissue. Twenty-eight clones showing overexpression in tumor versus normal tissue were then sequenced. These DNA sequences are provided in SEQ ID NO:63-92, respectively. The microarray analysis for these 28 clones is presented in
5 Table 7.

Table 7

| SEQ ID NO: | CLONE ID # | RATIO | MEDIAN SIGNAL 1 | MEDIAN SIGNAL 2 |
|------------|-------------------|-------|-----------------|-----------------|
| 63 | 72761 | 2.22 | 0.154 | 0.07 |
| 64 | 72762 | 2.33 | 0.105 | 0.045 |
| 65 | 72763 | 2.41 | 0.233 | 0.097 |
| 66 | 72764 | 2.72 | 0.199 | 0.073 |
| 67 | 72765 | 2.62 | 0.158 | 0.06 |
| 68 | 72766 | 2.84 | 0.149 | 0.053 |
| 69 | 72772 | 2.25 | 0.109 | 0.049 |
| 70 | 72775 | 2.36 | 0.103 | 0.044 |
| 71 | 72776 | 2.34 | 0.146 | 0.062 |
| 72 | 72779 | 2.25 | 0.22 | 0.098 |
| 73 | 72781 | 2.51 | 0.149 | 0.059 |
| 74 | 72784 | 2.35 | 0.212 | 0.09 |
| 75 | 72788 | 2.85 | 0.152 | 0.053 |
| 76 | 72789 | 2.69 | 0.196 | 0.073 |
| 77 | 72790 | 2.46 | 0.181 | 0.074 |
| 78 | 72791 | 2.39 | 0.143 | 0.06 |
| 79 | 72792 | 2.43 | 0.197 | 0.081 |
| 80 | 72794 | 3.04 | 0.258 | 0.085 |
| 81 | 72795 | 2.37 | 0.143 | 0.06 |
| 82 | 72797 | 2.96 | 0.233 | 0.079 |
| 83 | 72798 | 2.82 | 0.218 | 0.077 |
| 84 | 72804 | 2.33 | 0.14 | 0.06 |
| 85 | 72805 | 2.33 | 0.102 | 0.043 |
| 86 | 72806 | 2.32 | 0.121 | 0.052 |
| 87 | 72807 | 3.02 | 0.117 | 0.039 |
| 88 | 72808 | 2.74 | 0.109 | 0.04 |
| 89 | 72809 | 2.26 | 0.126 | 0.056 |
| 90 | 72811 | 2.92 | 0.151 | 0.052 |
| 91 | 72813 (L1080C) | 2.66 | 0.138 | 0.052 |

Each of the sequences was then used as a query to search the public sequence databases to identify novel and known genes. Results of this search are provided in Table 8.

5

Table 8

| SEQ ID NO: | GEN BANK ACC # | GENESEQ | DESCRIPTION |
|------------|---------------------------|----------------|---|
| 63 | AC004590 | | Chromosome 17 |
| 64 | Z78409 | T62661 | transcription factor E2F5 |
| 65 | S45828 | Z86797; A09328 | cDNA DKFZp564L2416; nekl=serine/threonine-and tyrosine-specific protein kinase [mice, erythroleukemia cells] |
| 66 | | | Novel |
| 67 | AL136169 | | Chromosome Xq26.1-27.1 |
| 68 | AC011742 AK021426 | | Chromosome 2, Homo sapiens cDNA FLJ11364 fis. clone HEMBA 1000264. |
| 69 | NM 005414 | Q03742 | SKI-like (SKIL) |
| 70 | NM 002335 | V85551 | low density lipoprotein receptor-related protein 5 |
| 71 | XM_004587 AB000520 | | Homo sapiens adaptor protein with pleckstrin homology and src homology 2 domains (APS), mRNA. Homo sapiens mRNA for APS, complete cds. |
| 72 | AK024119 | | cDNA FLJ14057 fis, clone HEMBB1000337. |
| 73 | U86338 | | Mus musculus zinc finger protein Png-1 (Png-1) |
| 74 | | | Novel |
| 75 | | | Novel |
| 76 | NM_002271 | C03734 | Homo sapiens karyopherin (importin) beta 3 (KPNB3) mRNA |
| 77 | NM_001401 | T48669; | Homo sapiens endothelial |

| | | | |
|----|-----------|-------------|---|
| | | T44104 | differentiation, lysophosphatidic acid G-protein-coupled receptor, 2(EDG2), mRNA. |
| 78 | U40583 | | Human alpha / neuronal nicotinic acetylcholine receptor mRNA, complete cds. |
| 79 | | Z15509 | Novel |
| 80 | Z59860 | V34162 | H. sapiens CpG island DNA genomic Msel fragment, clone 178c7, reverse read cpg178c7.rtla. |
| 81 | | | Novel |
| 82 | Z59860 | HNGIT2 2 | DNA genomic Msel fragment, clone 178c7 |
| 83 | XM-004477 | Q72451 | Homo sapiens glutamate-cysteine ligase, catalytic subunit (GCLC), mRNA. |
| 84 | | Z16421 | Novel |
| 85 | | | Novel |
| 86 | AC022013 | V52850 | Chromosome 3 |
| 87 | | | Novel |
| 88 | AL354993 | Z91766 | Chromosome 20q13.2-13. Continas a peptidylprolyl isomerase A (cyclophilin A) pseudogene, the gene for OVC10-2, ESTs, STSs and GSSs, complete sequence |
| 89 | AC005021 | | Chromosome 7q21-q22, complete sequence. |
| 90 | AK023904 | | cDNA FLJ13842 fis, clone THYRO1000793. |

EXAMPLE 9

QUANTITATIVE REAL-TIME PCR ANALYSIS

5

One of the clones of Example 7, clone L1080C, was further selected for Real-time PCR analysis. The Real-time PCR was carried out as disclosed in Example 3. The Real-time PCR shows over-expression in small cell lung carcinoma as well as in

brain and pituitary. Expression was also observed in thyroid, adrenal and salivary glands.

EXAMPLE 10

5 IDENTIFYING FULL-LENGTH cDNA SEQUENCE ENCODING L1080C

The cDNA sequence encoding full-length L1080C was predicted by using a partial sequence as a query to search the public sequence databases to obtain extended sequence. The query resulted in the identification of a full-length cDNA sequence for L1080C (SEQ ID NO:91). The deduced amino acid sequence encoded by 10 the full-length cDNA sequence is provided in SEQ ID NO:92.

EXAMPLE 11

PEPTIDE PRIMING OF T-HELPER LINES

Generation of CD4⁺ T helper lines and identification of peptide epitopes 15 derived from tumor-specific antigens that are capable of being recognized by CD4⁺ T cells in the context of HLA class II molecules, is carried out as follows:

Fifteen-mer peptides overlapping by 10 amino acids, derived from a tumor-specific antigen, are generated using standard procedures. Dendritic cells (DC) are derived from PBMC of a normal donor using GM-CSF and IL-4 by standard 20 protocols. CD4⁺ T cells are generated from the same donor as the DC using MACS beads (Miltenyi Biotec, Auburn, CA) and negative selection. DC are pulsed overnight with pools of the 15-mer peptides, with each peptide at a final concentration of 0.25 µg/ml. Pulsed DC are washed and plated at 1 x 10⁴ cells/well of 96-well V-bottom plates and purified CD4⁺ T cells are added at 1 x 10⁵/well. Cultures are supplemented 25 with 60 ng/ml IL-6 and 10 ng/ml IL-12 and incubated at 37°C. Cultures are restimulated as above on a weekly basis using DC generated and pulsed as above as antigen presenting cells, supplemented with 5 ng/ml IL-7 and 10 U/ml IL-2. Following 4 *in vitro* stimulation cycles, resulting CD4⁺ T cell lines (each line corresponding to one well) are tested for specific proliferation and cytokine production in response to the 30 stimulating pools of peptide with an irrelevant pool of peptides used as a control.

EXAMPLE 12

GENERATION OF TUMOR-SPECIFIC CTL LINES USING IN VITRO WHOLE-GENE PRIMING

Using *in vitro* whole-gene priming with tumor antigen-vaccinia infected

- 5 DC (see, for example, Yee et al, *The Journal of Immunology*, 157(9):4079-86, 1996),
human CTL lines are derived that specifically recognize autologous fibroblasts
transduced with a specific tumor antigen, as determined by interferon- γ ELISPOT
analysis. Specifically, dendritic cells (DC) are differentiated from monocyte cultures
derived from PBMC of normal human donors by growing for five days in RPMI
10 medium containing 10% human serum, 50 ng/ml human GM-CSF and 30 ng/ml human
IL-4. Following culture, DC are infected overnight with tumor antigen-recombinant
vaccinia virus at a multiplicity of infection (M.O.I) of five, and matured overnight by
the addition of 3 μ g/ml CD40 ligand. Virus is then inactivated by UV irradiation.
CD8+ T cells are isolated using a magnetic bead system, and priming cultures are
15 initiated using standard culture techniques. Cultures are restimulated every 7-10 days
using autologous primary fibroblasts retrovirally transduced with previously identified
tumor antigens. Following four stimulation cycles, CD8+ T cell lines are identified that
specifically produce interferon- γ when stimulated with tumor antigen-transduced
autologous fibroblasts. Using a panel of HLA-mismatched B-LCL lines transduced
20 with a vector expressing a tumor antigen, and measuring interferon- γ production by the
CTL lines in an ELISPOT assay, the HLA restriction of the CTL lines is determined.

EXAMPLE 13

GENERATION AND CHARACTERIZATION OF ANTI-TUMOR ANTIGEN MONOCLONAL

- 25 ANTIBODIES

Mouse monoclonal antibodies are raised against *E. coli* derived tumor
antigen proteins as follows: Mice are immunized with Complete Freund's Adjuvant
(CFA) containing 50 μ g recombinant tumor protein, followed by a subsequent
intraperitoneal boost with Incomplete Freund's Adjuvant (IFA) containing 10 μ g
30 recombinant protein. Three days prior to removal of the spleens, the mice are
immunized intravenously with approximately 50 μ g of soluble recombinant protein. The

spleen of a mouse with a positive titer to the tumor antigen is removed, and a single-cell suspension made and used for fusion to SP2/O myeloma cells to generate B cell hybridomas. The supernatants from the hybrid clones are tested by ELISA for specificity to recombinant tumor protein, and epitope mapped using peptides that span 5 the entire tumor protein sequence. The mAbs are also tested by flow cytometry for their ability to detect tumor protein on the surface of cells stably transfected with the cDNA encoding the tumor protein.

EXAMPLE 14

10

SYNTHESIS OF POLYPEPTIDES

Polypeptides are synthesized on a Perkin Elmer/Applied Biosystems Division 430A peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence is attached to the amino terminus of the peptide to provide a method 15 of conjugation, binding to an immobilized surface, or labeling of the peptide. Cleavage of the peptides from the solid support is carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides are precipitated in cold methyl-t-butyl-ether. The peptide pellets are then dissolved in water containing 0.1% trifluoroacetic acid (TFA) 20 and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0%-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) is used to elute the peptides. Following lyophilization of the pure fractions, the peptides are characterized using electrospray or other types of mass spectrometry and by amino acid analysis.

25

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS

What is Claimed:

1. An isolated polynucleotide comprising a sequence selected from the group consisting of:

- (a) sequences provided in SEQ ID NO:1-3, 5, 7, 9, 11-19, 25-35, 44, 46, 47, 48, 53-55, 58-60, 66, 74, 75, 79, 81, 84, 85, 87, 93, 94 and 95;
- (b) complements of the sequences provided in SEQ ID NO:1-3, 5, 7, 9, 11-19, 25-35, 44, 46, 47, 48, 53-55, 58-60, 66, 74, 75, 79, 81, 84, 85, 87, 93, 94 and 95;
- (c) sequences consisting of at least 20 contiguous residues of a sequence provided in SEQ ID NO:1-3, 5, 7, 9, 11-19, 25-35, 44, 46, 47, 48, 53-55, 58-60, 66, 74, 75, 79, 81, 84, 85, 87, 93, 94 and 95;
- (d) sequences that hybridize to a sequence provided in SEQ ID NO:1-3, 5, 7, 9, 11-19, 25-35, 44, 46, 47, 48, 53-55, 58-60, 66, 74, 75, 79, 81, 84, 85, 87, 93, 94 and 95, under highly stringent conditions;
- (e) sequences having at least 75% identity to a sequence of SEQ ID NO:1-3, 5, 7, 9, 11-19, 25-35, 44, 46, 47, 48, 53-55, 58-60, 66, 74, 75, 79, 81, 84, 85, 87, 93, 94 and 95;
- (f) sequences having at least 90% identity to a sequence of SEQ ID NO:1-3, 5, 7, 9, 11-19, 25-35, 44, 46, 47, 48, 53-55, 58-60, 66, 74, 75, 79, 81, 84, 85, 87, 93, 94 and 95; and
- (g) degenerate variants of a sequence provided in SEQ ID NO:1-3, 5, 7, 9, 11-19, 25-35, 44, 46, 47, 48, 53-55, 58-60, 66, 74, 75, 79, 81, 84, 85, 87, 93, 94 and 95.

2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

- (a) sequences having an amino acid sequence of any one of SEQ ID NO:61, 62 and 96;

- (b) sequences encoded by a polynucleotide of claim 1;
- (c) sequences having at least 70% identity to a sequence encoded by a polynucleotide of claim 1; and
- (d) sequences having at least 90% identity to a sequence encoded by a polynucleotide of claim 1.

3. An expression vector comprising a polynucleotide of claim 1 operably linked to an expression control sequence.

4. A host cell transformed or transfected with an expression vector according to claim 3.

5. An isolated antibody, or antigen-binding fragment thereof, that specifically binds to a polypeptide of claim 2.

6. A method for detecting the presence of a cancer in a patient, comprising the steps of:

- (a) obtaining a biological sample from the patient;
- (b) contacting the biological sample with a binding agent that binds to a polypeptide of claim 2;
- (c) detecting in the sample an amount of polypeptide that binds to the binding agent; and
- (d) comparing the amount of polypeptide to a predetermined cut-off value and therefrom determining the presence of a cancer in the patient.

7. A fusion protein comprising at least one polypeptide according to claim 2.

8. An oligonucleotide that hybridizes to a sequence recited in SEQ ID NO:1-3, 5, 7, 9, 11-19, 25-35, 44, 46, 47, 48, 53-55, 58-60, 66, 74, 75, 79, 81, 84, 85, 87, 93, 94 and 95 under highly stringent conditions.

9. A method for stimulating and/or expanding T cells specific for a tumor protein, comprising contacting T cells with at least one component selected from the group consisting of:

- (a) polypeptides according to claim 2;
- (b) polynucleotides according to claim 1; and
- (c) polynucleotides having a nucleotide sequence of any one of SEQ ID NO:4, 6, 8, 10, 20-24, 42, 43, 45, 49-52, 63-65, 67-73, 76-78, 80, 82, 83, 86 and 88-91;
- (d) antigen-presenting cells that express a polynucleotide according to claim 1,

under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells.

10. An isolated T cell population, comprising T cells prepared according to the method of claim 9.

11. A composition comprising a first component selected from the group consisting of physiologically acceptable carriers and immunostimulants, and a second component selected from the group consisting of:

- (a) polypeptides according to claim 2;
- (b) polynucleotides according to claim 1;
- (c) polynucleotides having a nucleotide sequence of any one of SEQ ID NO:4, 6, 8, 10, 20-24, 42, 43, 45, 49-52, 63-65, 67-73, 76-78, 80, 82, 83, 86 and 88-91;
- (d) antibodies according to claim 5;
- (e) fusion proteins according to claim 7;
- (f) T cell populations according to claim 10; and
- (g) antigen presenting cells that express a polypeptide according to claim 2.

12. A method for stimulating an immune response in a patient, comprising administering to the patient a composition of claim 11.

13. A method for the treatment of a lung cancer in a patient, comprising administering to the patient a composition of claim 11.

14. A method for determining the presence of a cancer in a patient, comprising the steps of:

- (a) obtaining a biological sample from the patient;
- (b) contacting the biological sample with an oligonucleotide according to claim 8;
- (c) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; and
- (d) compare the amount of polynucleotide that hybridizes to the oligonucleotide to a predetermined cut-off value, and therefrom determining the presence of the cancer in the patient.

15. A diagnostic kit comprising at least one oligonucleotide according to claim 8.

16. A diagnostic kit comprising at least one antibody according to claim 5 and a detection reagent, wherein the detection reagent comprises a reporter group.

17. A method for the treatment of lung cancer in a patient, comprising the steps of:

- (a) incubating CD4+ and/or CD8+ T cells isolated from a patient with at least one component selected from the group consisting of: (i) polypeptides according to claim 2; (ii) polynucleotides according to claim 1; and (iii) antigen presenting cells that express a polypeptide of claim 2, such that T cell proliferate;

(b) administering to the patient an effective amount of the proliferated T cells,
and thereby inhibiting the development of a cancer in the patient.

SEQUENCE LISTING

<110> Corixa Corporation
Harlocker, Susan L.
Wang, Tongtong
Bangur, Chaitanya S.
Klee, Jennifer
Switzer, Anne

<120> COMPOSITIONS AND METHODS FOR THE THERAPY
AND DIAGNOSIS OF LUNG CANCER.

<130> 210121.502PC

<140> PCT

<141> 2001-05-25

<160> 96

<210> i

<211> 644

<212> DNA

<213> Homo sapiens

<220>

<221> misc feature

<222> (1) (644)

6333 p = 3 T G ex G

5400> 1

<210> 2

<211> 1115

<212> DNA

<213> *Homo sapiens*

<400> 2

```

gttagaaagt acagtaaat gtatgttcat cttacttaca cacatagcta atctttttttt 60
tttcacttgtt aattatggc aaatgtttcat tggtacaaaaa aagttagacta gaaggatgtatg 120
ytcttaaaa gtcttgatc cattataaa gaaagaaaaa ggtaggagga agacagaaaa 180
gctgagactg gctgtatgc agggactta ctctccccc gggaaagactc ccggaccaac 240
attaacgaaa cagttttttg ttgtttttgtt tttttccca gcaatctta aagtgtatcc 300
catggacttc gccaaggaca ctttttaaag attatgtact ggcaatggca ttggccccgg 360
ggggccactc cttcacactg cttttttttt cttttttttt ctggggctca ttactttttttt 420

```

<210> 3

<211> 540

<212> DNA

<213> Homo sapiens

<400> 3

```

ggcccaagat tcggccgagg cttgc当地aaacg agaaggctgt ggattt当地gatt atttgtaccaa 60
gtgttc当地tctgtt aattt当地ata ctactt当地aaagc ttgtt当地cgat ggc当地acgtcc tcca当地acggcc 120
ttgtt当地aaaat ggttccctccg ttgtt当地acaa gaaaacggca aagaat当地ttac gaaggat当地tgty 180
atctt当地ttt aat当地tcaattt tgtaactt当地gtaca ttaat当地gttag tttagaaaatg tttagt当地ttta 240
actt当地aaaaaa aattt当地gttg ttgtt当地ccaa ttgtt当地tggt aaatc当地cggtt agtataccgtt 300
ggttt当地ttttt cc当地cccaagat aataaagggaa tagaca当地aacct cttt当地aaat tttt当地accat 360
taat当地ggaaaa aagt当地ttttt aat当地caatac aaat当地caaaaca attt当地aaatat ttttaa当地aaaa 420
aaggaaaaatg agat当地atgtg agt当地aggggaa aaaaat当地aaaaat gattc当地aaat tttatg当地tggaa 480
gaaaacccatg caat当地tttacc taga当地cqgtt ttaat当地atgtc tqttt当地ccca tctt当地atca 540

```

<210> 4

<211> 2076

<212> DNA

<213> Homo sapiens

<400> 4

| | | | | | | |
|--------------|--------------|-------------|--------------|--------------|-------------|------|
| agggtgccta | gctggccccc | gagcggttc | tccacctgg | gcagacacca | cctcggttg | 60 |
| catcgccgcg | ccccccctcg | gotggccct | aggccccc | cgtgtcctc | gcagccgcac | 120 |
| cccccccgca | gtgacagccg | ccggccggcc | tccggatcc | gatagtgtt | aagaaggaaa | 180 |
| tttacccctt | tctgtccaaa | tgaagtgtg | tttaatttgt | aaccatgtt | atccggact | 240 |
| taaaatcgta | aaacctgtat | acatggaaat | actatgttc | tacacccctt | catacttgg | 300 |
| aggttctgt | aaagactgca | ttaaaagacta | tgaaggctt | tcatgttgg | tgatccatcc | 360 |
| tgtgagccct | aggatgtac | aacttggaaa | tgaagcaag | cgcttgcata | acaaggaaaa | 420 |
| tcacaaatgt | caacacggc | ttaatagtac | aaatggaaa | gaagactgt | agaccatgg | 480 |
| attttatgaa | gacagtgcgt | atttcattt | ttcttcataa | atggccttca | tgatcaatcc | 540 |
| agaaggtagc | ctctggagg | agaatttcg | tgacacttta | caatcttgc | tgtcacaaat | 600 |
| acaaggccca | gaccaatatac | ccaaacaaaa | cttgcgtcc | gttccatcat | ttggaaaatg | 660 |
| gggttttgc | atcaaaaaaaaa | aaatgtccaa | acggaaatct | aaatgtatcc | ggggatgtat | 720 |
| gaaggaaaaatt | atagcccgag | gaattttttgc | atcgcaaat | ataattggca | gaaaatggg | 780 |
| ccttagaaatgt | tgatgatattc | tcagcgaat | ctttcgaaag | ggactccatgc | atgttcttag | 840 |
| acttattttt | gcaacactta | tgacatgg | cttaatcaat | tgttcaaaag | tgacaccaaa | 900 |
| ttggaaagaa | atccatggaa | atgtatggc | ggatccatcg | ttgtacatgt | agaocataac | 960 |
| aggatgttcc | gaaaacaaaca | ataaaatttc | acccatgtt | tcaacccat | atatgtttat | 1020 |
| gttcggaaacc | ccactggctt | ctgttcagaa | atccagcgc | cagacttcc | tcaaaaaaaa | 1080 |
| tgctcaaaacc | aaatgttccca | atccaatgg | tttttttttttt | tctactata | tgccggaaaa | 1140 |
| tttaatccctt | gagggttgc | acagatggaa | aaaaaaaccc | tgccggaaaa | ccgttgcattc | 1200 |
| ctgttaatca | ctgtccaaat | atgttgctt | tttacaaacgg | ggcaacccgtca | aacgagaaatg | 1260 |
| ctgtggattt | gattttatgg | ccaaatgttt | tttttttttttt | cataactacta | aaatgtttatc | 1320 |
| ataatggccaa | ctctccaaatg | ccatgtttat | tttttttttttt | cttccttgcgtt | aaaaaaaatgg | 1380 |

<210> 5

<211> 634

<212> DNA

<213> Homo sapiens

<400> 5

<210> 6

<211> 3725

<212> DNA

<213> Homo sapiens

<400> 6

```

accgttaaat ttgaaacttg ggggtatgg gtgtgggtt gaggtggccg gttttagg 60
gagtctgtg ctgttccttg tcgtttctgt agtcggagg gcaggttcg gaagaaggc 120
caggcgctc tgcccccaga gggaaaggcc gggcaggaccc gaggacttcg cccaaatgg 180
caggatggc ggcttggaga aagaaggggg tgcctgtat gagccatctg ctccggggg 240
agatgaatgg aactgttca aaaaatgtt acaaaccctta aggcaggccg ggatcatgc 300
cacgttcag ggacacttg cacaagaatc tgctgttaac aataatcttc acagcagaa 360
acggcattt gaaatgttcaa taggtatggaa tccgttttactgttgcggat gggctctgtt agttttgg 420
aacgttatta aatgttccatc taggtatggaa aacgttataa acttgcggat aacgttataa acttgcggat 480
tcgttatttc aatcttcgtt cgttacttgc taaatagg gggtttatgc aatgcgttcc ttggatgtatc 540
cgttacttgc caaccaacca aagaagatgg aactttagt ggatgttgtt ttcactgttgc cgttacttgc caaccaacca 600
tcaacagaag tggttcaacca aagaatgtt aacactttagt ggaaacqat qatggatattt tggatgtatc 660
tcaacagaag tggttcaacca aagaatgtt aacactttagt ggaaacqat qatggatattt tggatgtatc 720
agtgttcgg caaacttctgt ttgttacca caacgaacca aacactttagt ggaaacqat qatggatattt tggatgtatc 780
ttgttacca caacgaacca aacactttagt ggaaacqat qatggatattt tggatgtatc 840
agcttccatc atccgttgc aatcttcattt cttcaacacca aacactttagt ggaaacqat qatggatattt tggatgtatc 900
atcttcattt cttcaacacca aacactttagt ggaaacqat qatggatattt tggatgtatc 960
ttgttacca caacgaacca aacactttagt ggaaacqat qatggatattt tggatgtatc 1020
catggccagg gggaaacccaa atggatgttca agcaggccctt atggatgttca agcaggccctt tggacacatcg ccaggcttcc 1080
ggacacacagg ctccgttgcggc ttacttcattt aatcttcattt cttcaacacca aacactttagt ggaaacqat qatggatattt tggatgtatc 1140
tttacttcattt atgttggaa aactgttca aacactttagt ggaaacqat qatggatattt tggatgtatc 1200

```

acctagtata aaccacatcc taagcaccag aaaggccttga aaggaagaag gagatcctct 1320
 acaaagggtt cagagccatc agcaagcatc tgaggagaag aaagagaaga tgatgtattg 1380
 taaggaaag atttatgcag gagtagggga atttccttt gaagaaatcc gggctgaatgt 1440
 ttccggaaag aaaaaaaaaa agcaaaagggtt agcccgacta ttgaccatgt cagagaagag 1500
 agcagaatgt cagaaaacaga ttgaagagat ggagaagaag ctaaaagaaaa tccaaactac 1560
 tcagcaagaa agaacaagggtt atcagcaaga agagacgtg cctacaagg agacaactaa 1620
 actgcaattt gttcccgatc ctccaaaaat acccagaatgt actctatcca ttctgtttt 1680
 tcaagtaaac tggtgtccca gagaacactt acttgcggag aacatttggc aggaacaacc 1740
 tcattttttaa ggtcccgatc taccccccattttgtat gagtttottt ttccagaaaa 1800
 gaagaataaa agtcctccgtt cagatcccccc acggattttta gctcaacggaa gacccttgc 1860
 agttcccaaa accttcggaaat gcatccatc aaatgtatggatgttcccgatgttttgta 1920
 tgaattttca ggaattttccatc ctttgcggatc ggatccattt atacacggctt tcagaaatgt 1980
 aacaattttgtt cctaaccatc aagacacttgc tgacttttgc agagcagctc gttttgtatc 2040
 cactccctt catgatggatc tgcttgcggatgttcccttgc tgatgtctggatgttcccttgc 2100
 acccgaaagaa gtccatgtt gaaacacttgc tgaggacccatc cagacacgtt ttgcactat 2160
 ctacatcgatc acttccatc tcaagaacgtt gacccatattt attgaagaca gtcgtgaac 2220
 cacacactcc ttggccatctc ttggccatctc tgccctcggtt gcaagcacctt cttccatcaa 2280
 atgttccatc atttccatc aacttacactt tactaaatgtt acttccatgttccatc 2340
 gtccatcgatc tggttcacatc acatccatc acatccatc tgactttgc tgccatccatc 2400
 ctgttcggatc ttgtgtatgc aagacacacttgc aatgccttgc ttggaaatttgc aagaaatgtt 2460
 tgaattttgtt aatggatgtt actgcattttaa acggatataatc ctaatatgtt aatggatataa 2520
 gtatgttttttggccatc gaaacacttgc aatggatataatc gtaataaaggg tattttttccatc 2580
 acctgtccca ttggacttttca atcaatccatc caatggatataatc gaaacgtttaa atgaagattt 2640
 tgatcttttccatc atcaatccatc agatggatgtt attgttttgc accaatatataatc 2700
 aaatgttccca acccttcggatc atttccatc acatccatc tatattaccatc atgaatataac 2760
 agtgttgcattt atttataatc ttggatataatc agtgttgcattt ctacacaaatg cagaatataatc 2820
 ccatccatc ttggatgttccatc ggttgcgtat tttccatc aacatccatc agatccacg atccctatgc 2880
 ttgttccatc aacatccatc ttggatataatc agtgttgcattt tttccatc aacatccatc ttggatgttccatc 2940
 ggttgcgtatc gatgtttttccatc ttggatgttccatc gtttccatc aacatccatc ttggatgttccatc 3000
 aaatgttccca gtttccatc ttggatgttccatc gtttccatc aacatccatc ttggatgttccatc 3060
 agcaccatccatc ttatgttccatc agggatgttccatc ttggatgttccatc ttggatgttccatc 3120
 acttggatgttccatc aatatttttgc agttaaaatgc ttggatgttccatc ttggatgttccatc 3180
 gatgttgcattt ttggatgttccatc agggatgttccatc ttggatgttccatc ttggatgttccatc 3240
 gaatggatgttccatc aatccatc aacatccatc ttggatgttccatc ttggatgttccatc 3300
 gaatggatgttccatc aatccatc aacatccatc ttggatgttccatc ttggatgttccatc 3360
 ttgttccatc agggatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc 3420
 atttggatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc 3480
 tgactgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc 3540
 tgaatggatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc 3600
 ttgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc 3660
 tcacccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc 3720
 aaaaaaaatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc ttggatgttccatc 3725

<210> 7
 <211> 567
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)...(567)
 <223> n = A,T,C or G

<400> 7
 ggccaagaat tggccacccatc gacaacatc taaagggccatc aggccatgc ttgttgcggatc 60
 ttctccatccgg gggaaaaacccatc actgttgcggatc ttgttgcggatc ttgttgcggatc 120
 gctccggatc ttgttgcggatc ttgttgcggatc ttgttgcggatc ttgttgcggatc 180
 ttgttgcggatc ttgttgcggatc ttgttgcggatc ttgttgcggatc ttgttgcggatc 240

5

<210> 8

<211> 1365

<212> DNA

<213> *Homo sapiens*

<400> 8

<210> 9

<211> 1196

<212> DNA

<213> *Homo sapiens*

<400> 9

ataaattaaa attaaaaata tagagggaa tccttcgtat tgctcggcat gatgttagat 960
ataaggttt tcgcggatc atccatgtac gctgtttaac aaggatttt attaccatcc 1020
ttctaaaggcc atatggata ttgttatttg agagttcaac ctaataaagt tatccgttc 1080
ctttagacct tcttcgttaaa tagtgaaat ttggatggatc gatgtttttt aatccatcc 1140
aqqaaaaaaa aatggacqca ttgttcgtttaa ccatttttttccatggatc gatgtttttt 1196

<210> 10

£211> 1424

<212> DNE

<212> DNA

<400> 10

```

ctcagtcata gggaaatgaa gggtgtttt ctggcgata ctgaaaataga ccttttctc 60
acagacatcc ctccatcaa ccgcgtggac ttcaactggaa gatgttcatc cccaaaaatc 120
aaatataatgta acatggaa catcgatgt tttaaaatggg gaaacatgtt agtttacatc 180
gtgattgttc catgttgttcc ctgtttttt tctgcacaaca acagacactt ctggatgtt 240
cacaggccagg cagtttgc tatataacaga ctactacgtt cgggtttaa cgttccactt 300
ccccggcaact tggccagatg aagaagatgtt acatgttcaag atgtgtttaa tatctcgatc 360
gagggtgtta tttagataat ggaattatgat tataatgtat acacaaactt ttttttatc 420
aaaaataatataat taatgtatca actttttttt tttttttttt tttttttttt 480
caaaaatgggg goatttgtt attttttttt ttccgttctc taatcttact tttttttttt 540
atttgaccca ttggatgtt gttttttttt tttttttttt tttttttttt 600
ttcccaatgtt gatgtttttt tttttttttt tttttttttt tttttttttt 660
taccccccac tctggggggg tttttttttt aacttggatc cagatccatc taattttaca 720
ttcttgcata aaggccatgtt accacccgtt aacttcaacc tttttttttt aagtccacy 780
gaggttttaa aagggtttttt aaaaacaaacgg tttttttttt cttttttttt 840
aggttttttttt gtgtataactt gttttttttt ccggacaaaggg tttttttttt 900
ataatataatataat tttttttttt tttttttttt tttttttttt 960
aatatggatttt gtcagaaaaat atcgttatacc tttttttttt tttttttttt 1020
ttcttacccggc atttatggaa ttgtttttttt acatgttacac tttttttttt 1080
cttggggactt tttttttttt tttttttttt tttttttttt tttttttttt 1140
aaaaaaaaaaa actaaaggca ttggcttaa ccattgtaca tttaaaaggg ctgtttttttt 1200
ttgtttttttt ttgtttttttt tttttttttt tttttttttt tttttttttt 1260
ggaaatgttagg gtatgtttttt attctgtttttt tttttttttt tttttttttt 1320
aatatcttgcg ttgtttttttt ttgtttcaata cccaaaggatc tttttttttt 1380
taatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt 1424

```

<210> 11

<211> 460

<212> DNA

<212> BRI

<220>

<221> misc feature

<222> (1)

<223> n = A-T-C or G

<400> 11

```

agacgngac gatggaaaa gntcttaaca gatnattta atgaccctca gggtcgaat 60
gatnathgac ccagtggcac ntgggcattc tatggggca ntttgggttg gaacctcgat 120
ataaaatggc caaggcccaa caggatcatc ncggcttagag naganacna agatgtatg 180
gacgatgcg atagcngatg tggttgaat ngaaatttca gagaccccccc tcttacccgt 240
ttataaactatc caatgtatcg tggtccggc ggacaaacgat tggctactgc ttgcgtatg 300
cttccaagaat ttganaatcg ctcccccgtt atttgcgtc antttccaa acgngaaat 360
cttccaaatc tgccatggcc aattttatcg tggtttttcg genatgtctc tgntctcttg 420
cttccaaatc ctcccaaaaccc tttccacccctt gaagaaatgaa 460

```

<210> 12

```

<211> 2206
<212> DNA
<213> Homo sapiens

<400> 12
cagaacagac atgtgtgtg tgcagaccaa gaaggaggatt gccaggctgc ctccctgtc 60
cagaataaca cccgaaacag ccagaagcca tcggggaaa gactgtgc aaccacaaac 120
ttgtatccc aggagtccag ggggggattt caactgtcac gggaaatacta ctgtggaaaaat 180
ggccatggca agttgtactgt cggggatcc agaaacggcc ogggccccag ttccggactat 240
gtatgttcac cagccaaacg ggacccaaag cccttcgacc gtgtcgacca actgtcatca 300
gcctcccaat cccacacatg gocatgtcga atgtcccccc agggacccac ccacgttcgc 360
cctatggccg cggaaacacg gagacttatt gtcatgtaaa acgtgtggca gacccttcg 420
caggccggcc ccaggcttgg atgtggaaa gtgtgtccgt actgttttgc gaaacaaatg 480
ttgtatgtaa atcatggca caacgcgggt atgttgcggcc tgcatgtaaa ttgtgttgg 540
ggctggctca acatgttgcg acacccctt gaaatgtggc ctgtatgtca ctgtgtgtcc 600
caggatggaa ccaggccctt gcacgtgtcgt ttggaaaaac actatccgtt aatgtccgg 660
ctatcttcct ttatgttgcg tggcccccc ttgttgcgtt acttgttagt aacatcatgtg 720
aaaatgtaccc acatgtgtact ttatggaaa ttgtttaaacag attttttaaa tgccctcc 780
ggtcgtccaa atgtatgtcc cttgtggact tyggactttt atggacgtc tyttgttgg 840
ccatgtatgtt aaatgggtca ttgtttttt gccaacccccc caggccggcc agacccaggat 900
gtatgtatgtc acgttccatgtt ccgtatgtttt gaaatggat ttgttgcggcc cccttcctt 960
ccgttataca atcatccaaatgt atctgtgttccatgtt ccgtgggttgc catgtgttc atgtgttgg 1020
aaactatgtc acgttccatgtt gaggcccaacg aactgtgtcc tggttccgg tggccatgtt 1080
aaatgtaaaa ttgttccccc catatgttgc tggaaattttccaaatgttgc aatgtccacc 1140
atgtccatgtt ccgttccatgtt tggccgggtt tggatgtttt tggatgttccatgtt 1200
gaccgtggaa ctttcaacc tggaaatgtt gggccggcc tggatgttccatgtt 1260
gaaatgttca ctttgcgtgg ccgttccatgtt gggatgttccatgtt 1320
gaaacaaatgt tggatgttccatgtt ccgtggccccc catatgttgc tggatgttccatgtt 1380
gtgcataatgt gtccatataat aactatgttgc tggatgttccatgtt 1440
tatcttcctt ttatgttgc tggatgttccatgtt 1500
tccctttaaa ttgttgcgtt acgttccatgtt aacttggatgttccatgtt 1560
tgccacatcatgtt ccgttccatgtt tggatgttccatgtt 1620
aggagaaaaaaa atttttttttt gggatgttccatgtt 1680
ttgtatgtca ggcacatgtt gggatgttccatgtt 1740
attccggggaa ggcacatgtt gggatgttccatgtt 1800
aggatgttccatgtt tttttttttt gggatgttccatgtt 1860
cacaacaaatgtt aacttggatgttccatgtt 1920
cggccatgttccatgtt acgttccatgtt 1980
ttgtatgttccatgtt aacttggatgttccatgtt 2040
ttgttgcgttccatgtt aacttggatgttccatgtt 2100
ttgttgcgttccatgtt aacttggatgttccatgtt 2160
aaaggccatgttccatgtt aaaaaggaaaaaaa aaaaaaaaaaaaaaa 2206

```

<210> 13
<211> 680
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<222> (1)...(680)
<223> n = A,T,C or G

<210> 14

<211> 5023

<212>

<210> 15

<211> 403

<212> DNA

<213> Homo sapiens

<400> 15

```

ccatcacggg gaattctgtc gtgttatta cccattcaa gttgacaact gaggcaacgc 60
agacttcagg ctccaataag aaaaacctgt ttgtatcaa agcaatggt ttcgtcccc 120
ttaactatc accacacaaa ggaaaacgtt aaccatgggg gcaatctaa gaaaataat 180
atctaatac atcatgttcaaa aaaaaatcaa ttcaacaaacg aacttacaa caaccacatc 240
ttccagacaa ggaagacgca cggaaacgaa ggacgaaaga acggaaaggaa aagaacccatc 300
aggttttggg atcgaaacgg ggccttattt tggtcttggaa ttatattt ttaaacatc 360
tttaataatc ctgttatcc tcactttt cttttttttt attttttttt attttttttt 403

```

<210> 16

<211> 890

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(890)
<223> n = A,T,C or G

<400> 16
agcataaagg tntcaactgac caagactcca gccagaaaat ctgcacatgt gaccgtgtct 60
ggggccaccc aaaaaggcga ggctgtgtc gggacacaca aattaaagac catcacgggg 120
aatttctgtc ctgttattac cccatccaq ttgacaactg aggcaacgcg aactccagtc 180
tccaaataaga aaccatgtgt tgatcttaaa gcaatgttgt ctgtccccc caactatgaa 240
ccacacaacaa gaaaggtaaa accatgggg caatctaag aaaaataatta tctaaatcaa 300
catgtcaaca gaattaaacct ctacaagaaa acttacaaac acccccatct ccagacaag 360
gaagagcaac ggaagaaaacg cgacaaaggaa cgaaaggaga agaaagcaaa ggttttggaa 420
atgogacagg ggcttcattt ggctgaatg taataatttt ttacatctt gtaaaatattc 480
ctgttatttc aacttttttc cttttgtaaa tttttttttt ttgtctgtca tccccactt 540
agtacacgaga ttctttttt ctaactgtt atagtctgtg gtatgttca tgggttctc 600
atgtgtatg atctctgtaaa agacgttatac acctttaaaac tcaaattttt tgggatgttt 660
tttacttaag tccattaaaca attcaggttt ctaacgagac,ccatccaaa attctgtt 720
tagattttt atgtcaattt cccaaagttyt ccctgctgg tctaattattha acagaactgc 780
agtotttgc tagccaaatg catttacgt atggcagcta gttatgccag cttagggag 840
aatttgaata ttttccagga atggggaaag ctgggaaaga aaggccacct 890

<210> 17
<211> 371
<212> DNA
<213> Homo sapiens

<400> 17
ttggctcagg aggacaatat ggtggaaat gacaaagtaa ctccctgtggc cctaggtcag 60
gttcttcttga ggaaaaaaa aaggctggaa tgaaatcact ctccgtaaac cagggtccctc 120
cagtgcgtgc ggttatttccc aagtccacat tttgcagaca gggccttaaa atgtctatgt 180
aggaagtcc tgagcctgtt tttttaaaat tctacacaca cacatgcaca cacacacgca 240
cgtgtgcaca catggggata tatacatctt caccctttct tgagattact gctcagaa 300
aggacacatt ggtttggctc gtttaccagg tgctgaatgt ggagccggcg caagcttawt 360
tccttttagt g 371

<210> 18
<211> 376
<212> DNA
<213> Homo sapiens

<400> 18
attctttggc tcagcaggac aatatggtg gaaatgacaa agtaactctt gtggccctag 60
gtcagggttctt ctgaggaaa acaaaaaggc tggaaatgata cagcttcttc taaaaccagg 120
gcctccatgt ctgcggtaa ttcccaactg cacattttc agacaggccc ctaaaatgtc 180
tagcttaggaa gttctgtggc ctgtttttttaaaatctac acacacacat gcacacacac 240
acgcacatgt acacacatgc ggtatataac atccctccactt tttcttgaga ttactgttca 300
gaagaaggcga cattttgtttt ggtctgttca ccagggtgtc aagtggggacg ggccgcacg 360
ttawttccctt ttatgtg 376

<210> 19
<211> 512
<212> DNA
<213> Homo sapiens

tctttgtt accagtgggt tttgtgaggat gtcagtagt aaaaactggc ctttacactga 2220
 cagagccctg gctttgacact gtcagccct gtgtgttaat cctctagtag ccaattaact 2280
 actctgggtt ggcagggtcc agagaatcga gtgacactt tgccactat ctgtgttta 2340
 cttagagacat gtaaaatatga tagggaagga actgaaatttcc tccatccata ttataaacc 2400
 ttctagttt atcttccttg gctttaagag tttgcccattg aaagtgtaaa gaaatgaaat 2460
 tcttaggtcaa gcaaaaagat gctggagata ttgtatacttcc tcattttaaac ttgtgtttta 2520
 tttacatcgat atgtactaaa ataagaataa tagaaatttt cttgttaggtt aatccatgt 2580
 agccaataat tttaaaatgtt ctttatctgtt atcatgtgtt ttgtgtacta taaaataaaat 2640
 gacccatcgat gaaagggttga ggttgtatcc ttgtgtttaat tttatgtt ggttccatgtt 2700
 gctacaataat atccagacta ccaggcttgg tagatattaa agctgggtac taagaatgt 2760
 tatttgcata ctctcgat tccctgttgc ttgttgcattt atacgttccccccaggagcatg 2820
 ctgttttgc aatacaataat aaatattatg gaggcttccccccaggatgttataat 2880
 gattgtctt ctctttataa taagagaaac aaattcttataat ttgtgtttaat aacatgttt 2940
 tttagtggat ctatgttgc tttttttttgc ttgtgttgc acgtgtgtaa aagtgttgc 3000
 ttgttatttttgc atatgttgc ttatgttgc gttttgtcgg ggtgttttgc 3060
 agtgccttc atcagggttga aggagatataat ttacggggatc ttgtgttgc 3120
 atgttactatg ttgttgc ttgttgc ttacaagaga ttgttttttgc 3180
 tgaatggggaa aataatccctt taaaatatgaa cggacatccaa cttagatgg gtttggggat 3240
 ttccaaacgc ttttttttttgc ttacaatgaa atgtgttgc ttgtgttgc 3300
 atatcttataat catagatcgat gtttttttttgc ttacaatgaa atttttaat 3360
 tcatttttttgc ttgtgttgc tttttttttgc ttacaatgaa aagatatttttgc 3410

<210> 21
<211> 627
<212> DNA
<213> Homo sapiens

<400> 21
 ggccaagaat tggcccgagg ggtggccggg ccatggagaa gcttagctcc atcaaatctc 60
 aaacaatttt tggatattttt gataatcttcc aaggatctca ctgtttgtccatggccccc 120
 aaaatagaaat caaatgttcaat attccatccccc gatgttccatggccatgtt 180
 tagaaaaaaat atttcttgc ttttttttttgc ttatgttgc cacaatgttgc gacgttccatggcc 240
 ggtctgtgggg aggccatccgg gccttcttgc ttatgttgc cacaatgttgc gacgttccatggcc 300
 agtgcggccccc ctatgttgc ttatgttgc ttatgttgc gatgttgc cacaatgttgc gacgttccatggcc 360
 agatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 420
 aaatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 480
 agaacaacacaaatccatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 540
 tctgttttttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 600
 ttgtgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 627

<210> 22
<211> 1065
<212> DNA
<213> Homo sapiens

<400> 22
 ctttggcttgc ttccacccccc tggccggccgc accatggaccc cccccaggca ggtgggtcaac 60
 ttggccctgttgc tggccggccaa gctggcccac tcgtgtgttgc tagatgttgc aaaggatattt 120
 tttagactaca aaggatgttttgc ctttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 180
 gccaagatataat ttttttttttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 240
 tataatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 300
 ctatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 360
 aaggcccgccatgttgc aaggatgttttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 420
 aaggcccgccatgttgc aaggatgttttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 480
 ttttttttttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 540
 ggacgttccatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 600
 aaggatgttttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 660
 gtgtgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc ttatgttgc 720

```
tacaagggtgc aggctggaaa cagctcccttg tacaacacgc ctccatgtt cagcatctac 780
gtcattgggt tggtttcttgg gtggataaa aacaatggg gtcgcggcc catggagaac 840
cttagttcca taaaatctca acaaattttt gagattttg ataatttctca agaggatctac 900
gtgtctgtgg gaggcatccg ggccctctctg tataatgtctg tcacaatttg aagaaccttcag 960
aaatggccgg ctttcattttt aaaaatttttt gagatgtttc agactatggaa acatcttaac 1020
caggatatac ttgtttcttgg aacaacatca aagaattttaa glaac 1065
```

<210> 23
<211> 578
<212> DNA
<213> *Homo sapiens*

```
<220>
<221> misc_feature
<222> (1)...(578)
<223> n = A, T, C or G
```

<210> 24
<211> 3799
<212> DNA
<213> Homo s

<400> 24
atagtaaaccc agaacttcaa atcctatgcg ggggagaaaa ttctgggacc ttccataag 60
cgcttttcgt ttttatccgg gccaaatggc agtgccaat ccaatgttat tgattctatg 120
ctttttgtgt tttggctatcg aacccaaaaaa ataaatgttca aaaaactctc agatataaa 180
caataatctg atgaacacaa ggacattcg agttgtacag tagaaatgttca ttttccaaa 240
ataattgtat aaggaaaggaa tgattatga gtcattccca acagtaattt ctatgtatcc 300
aaagaacggctt gaaagatcttctc tacttcgtc tatcacatgg aaagacattt 360
aaggatgtgtt gaaatcttctc tcgaacccat ggaattgtact tgccatataa tagatttta 420
attttacagg gtgaagggtt acaaaatgttctc atgatggaaac caaaaggccca gactgaacac 480
gtatgggtttt tgcttgcataa tttagaaatgttca ataaatgttgc tgccatggctt aaatggacat 540
ataaaatgtt tggtcaaaag agttgttgcataa tttaaaatgttca acagagggaa gaagttaaac 600
agggtttttttt tggtgtttttt gggaaaaggat gccttgcataa gagaaaaaaa catagtatcc 660
gaatttttttt cttttttttt tgaaatattt aaaaaaaaaaa atcatgtttt tcaatattttt 720
attttatgtt tgccaaaaagg attttgttgcataa atggggaaatc aaaaaaaaaaa aattttatgtt 780
gatccaaaaaa aattttatgttca ggaaaggccat atactatccat gttttttttt aatggaaatg 840
aaatgtttttt aatggatccat aaaaaaaaaaa aatggaaatc aaaaaaaaaaa tttttttttt 900
aaagaaaaat tttcacacatgtt agatttggaa gatgttgcataa tttagaaaaaa gttttttttt 960
ggccacatgtt aagccaaaaat actggggaaatc caacttccat aaaaaaaaaaa aatggatggaaat 1020
gaatttttttt tttatccatgttca ccagatggaa aatattttatgttca aatggaaacat 1080
aatggccctcg aagaaggaaaaa agagaaaaaa gaaaaaaaat taaaaggatgtt tttttttttt 1140
ctttaaaacagg aacacaaacagg gttttttttt gttttttttt gttttttttt gttttttttt 1200
gtttttttttt aatctggatgttca taatctgttca aatggatggggatgttccat gttttttttt 1260
gatataatctat ttcgtatgttca taatctgttca aatggatggggatgttccat gttttttttt 1320
caatggatgttca tttttttttt tttttttttt aatggatggggatgttccat gttttttttt 1380

<210> 25

<211> 429

<212> DNA

<213> Homo sapiens

<220>

<221> misc feature

<222> (1), (429)

<223> n = A, T, C or G

<400> 25

gttcctgggg ataatacatt aaccaacatt gggttggaaaca tacctgagta atcatatcag 360
 gatgcgttt aagctgataa aacaataaga tcccaaataat cagtagctca aaaaaaaaaa 420
 aaaaaaagn 429

<210> 26
<211> 788
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(788)
<223> n = A,T,C or G

<400> 26
 ncctttttttttttttttt gaggctactgc atttttggat cttaattttttt tatcagctta 60
 acatgcattttt tgatattttt actcaggat ttttcaacca atgttggta atgttatttc 120
 cccaggaaact tattttactaga ggagcgttg agtggtaat ttctggaaatg tggaaatttc 180
 attttttttt tttttttttt gacagtgcatt cttttttttt cttttttttt cttttttttt 240
 tacatataata tacattttttttt acctttttttt gttaatgtttt gtatgtttttt aactttttttt 300
 gaggaggatttt tttttttttt tgccattttt aaattttttt tttttttttt tttttttttt 360
 ctgtttttttt ttttttttt taatggatgg atgatgtttt ataatttttttaa aatactttttt 420
 tttttttttt tttttttttt tagccccccc tttttttttt tttttttttt tttttttttt tgatgtttttt 480
 agttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 540
 attttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 600
 agttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 660
 attttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 720
 agatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 780
 tagggaaaaa 788

<210> 27
<211> 687
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(687)
<223> n = A,T,C or G

<400> 27
 acatgtttttt tgctttttttt ttaaacatct ttaaaatgtt attttttttt atctgtttttt 60
 tgatgtttttt tttttttttt gataatccctt gatgtttttt gatgtttttt tgatgtttttt 120
 tttttttttt tttttttttt ataaatctttt ataaatctttt tttttttttt tttttttttt 180
 gtatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 240
 acatattttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 300
 taaaatccctt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 360
 taggtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 420
 gaggctttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 480
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 540
 agttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 600
 ngtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 660
 atacatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt 687

<210> 28
<211> 1529
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(527)
<223> n = A,T,C or G

<400> 32
atccaggag aggagtctat ctcccaagn ttgacaactc ctacttgg tggcggncaa 60
aatcagtcta ctacagatgc tattatacta gataaaaaat tngtgacaaa gtctggagtc 120
tagggttggg cagaagatga catttaattt ggaaattttct ttttactttt gtggagcatt 180
agactcacag ttatccat tgatattttt ctgtatggntt gtgaactctt gctggaaatc 240
aaaattttct tgagactt tagcatctat tttttgggnn taaaaggat tntctcagact 300
catccagccc ttgggtgtg accagcagag tcaactagngg atgctgaagt tacatgagct 360
acatgttaaa tattttaaatg ctccaaaataa aaacccccca acgttgaccc taccggcgtt 420
gtatgttagt cccttgggtg gctgtccat gtgcctttagt agagcccgta agttacagtt 480
gtctctaat ttgaaatcca taagntaaaca ngtctatatac agntgcn 527

<210> 33
<211> 934
<212> DNA
<213> Homo sapiens

<400> 33
gtaggccagc gatgacgacg aggaggaaga agggaaacatc ggttgtgaag agaaagccaa 60
aaagaatgcc aacaaggcctt tgctggatg gattgtgcctt gtgtccggacg ggactgtcat 120
gaggatgtt atgctggcag ccatcaataat ccaaggagaa ggagttttatc tcctcaagtt 180
tgacaatcc tctttttgtt ggccgtcaaa atcagtctac taatggatctt attataactatg 240
ataaaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 300
gaaattttctt ttactttttt tttttttttt tttttttttt tttttttttt tttttttttt 360
tgatgtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 420
ctttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 480
caactatgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 540
aacaccccaaa cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 600
tctttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 660
atataatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 720
cacaatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 780
gattttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 840
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 900
tccaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 934

<210> 34
<211> 758
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(758)
<223> n = A,T,C or G

<400> 34
ggctttatag cccatcctca ttgtttactg ccacccctca gctggggtcc aaggcagttac 60
tattttttttt attttttttt attttttttt attttttttt attttttttt attttttttt 120
tccatcaagg agcatgtttcc agagttttcc ccagatgttcc cccaaagaaac actgtccgg 180
gctgtggcgtt atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 240
gaatccgtttt atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 300
acacaaatctg atttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 360
tttaaagctgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 420

19

gaacgaact gagtttccaa gggtttgggg ctgtgcaaga gtaaacacta gagcttgagt 480
 tttatccag ctggcaagca cgaaagtct tgaagaatgt aatgtaaaaa gggaaaagaa 540
 tgtaaagctt ttgtaccaa attagagttt gggccaccc aacaaatgtt ttccctgtg 600
 taaaatgtct tctggaggc acatccatc tccatggtc actctgaggc gcactgtca 660
 cttagatgtt gccccatcca ggtggaggc acccttttg gatggngagt atncaatctg 720
 ctngcattt tgacaggata tctgaatggc tagtaat 758

<210> 35

<211> 1534

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)...(1534)

<223> n = A,T,C or G

<400> 35

ngaggtaaaa ggcaaggcag catttaataa stacctgtt tattttttt aatgtttttt 60
 gtggtaatcc tcacaagac cggactgtt gaaactctt tgcttattaa ttttttttt 120
 tgaggaaattt tgcttttccaa gtgcataac acttataata tttttttccaa aagaggagca 180
 ttctggatcta atttatgtcc ttgtgactgtt ttaaggatctt agcttcttc agagctggc 240
 tatccggatgtt gatctggcc ttctctacaga ttgttctgtt atgcctggcc aaggatggcc 300
 cttagccaggta taacagctttt atagcccatc ttcatgtt actgcaccc ctcagctgg 360
 gtccaaaggca gtactattca ttccatgtt cagacccatc tccagacatc tactttttc 420
 aaaaatttagt ttttttccatc aaggagatgtt ttccagacca ttcccccagat atgccc 480
 gaacactgtc cgtgtcgtt ggttacatgtt gcaacacgtt tagactaaatg ggaacatccc 540
 agcaggctgtt tttaatggc ttcttattttt ttatgttacttca tttttttttt 600
 aaacgcggca cacacacaca atctgtatggc goataatcttca tttttttttt 660
 ttcttcttctt ctgttaatggc ttatgttacttca tttttttttt 720
 ggtttagtgc actgtatggaa cgaatgttacttca tttttttttt 780
 aaacactaga gttttagtgc ttatccatgtt gcaacacgtt gaaatgttacttca 840
 tgtaaaaagg gaaaaggatgtt taaatgttacttca tttttttttt 900
 caaatgttttccctgttca aaatgttcttca tttttttttt 960
 ttctggggc actgttcaactt agatgttacttca tttttttttt 1020
 ttgttagatgtt ccaatgttacttca tttttttttt 1080
 ccaaggccggc ttccatgttacttca tttttttttt 1140
 ttatccatgttacttca tttttttttt 1200
 aatatttttttccctgttacttca tttttttttt 1260
 agccatggc accccatccca aattttccatgttacttca tttttttttt 1320
 ttctgttacttca aagggttcttc tttttttttt 1380
 aatatttttttccctgttacttca tttttttttt 1440
 atgttagtgc caaatttttttccatgttacttca tttttttttt 1500
 tttttaacaa ataaacttacatgttacttca tttttttttt 1534

<210> 36

<211> 125

<212> PRT

<213> Homo sapiens

<220>

<221> variant

<222> (1)...(125)

<223> Xaa = Any amino acid

<400> 36

Leu Ser Ser Arg Gly Met Lys Ala Val Leu Leu Ala Asp Thr Glu Ile

Asp Leu Phe Ser Thr Asp Ile Pro Pro Thr Asn Ala Val Asp Phe Thr
20 25 30

Gly Arg Cys Tyr Phe Thr Lys Ile Cys Lys Cys Lys Leu Lys Asp Ile
35 40 45

Ala Cys Leu Lys Cys Gly Asn Ile Val Xaa Tyr His Val Ile Val Pro
50 55 60

Cys Ser Ser Cys Leu Leu Ser Cys Asn Asn Arg His Phe Trp Met Phe
65 70 75 80

His Ser Gln Ala Val Tyr Asp Ile Asn Arg Leu Asp Ser Thr Gly Val
85 90 95

Asn Val Leu Leu Arg Gly Asn Leu Pro Glu Ile Glu Glu Ser Thr Asp
100 105 110

Glu Asp Val Leu Asn Ile Ser Ala Glu Glu Cys Ile Arg
115 120 125

<210> 37

<211> 448

<212> PRT

<213> Homo sapiens

<220>

<221> VARIANT

<222> (1)...(448)

<223> Xaa = any amino acid

<400> 37

Met Ser Arg Arg Pro Cys Ser Cys Ala Leu Arg Pro Pro Arg Cys Ser
5 10 15

Cys Ser Ala Ser Pro Ser Ala Val Thr Ala Ala Gly Arg Pro Arg Pro
20 25 30

Ser Asp Ser Cys Lys Glu Glu Ser Ser Thr Leu Ser Val Lys Met Lys
35 40 45

Cys Asp Phe Asn Cys Asn His Val His Ser Gly Leu Lys Leu Val Lys
50 55 60

Pro Asp Asp Ile Gly Arg Leu Val Ser Tyr Thr Pro Ala Tyr Leu Glu
65 70 75 80

Gly Ser Cys Lys Asp Cys Ile Lys Asp Tyr Glu Arg Leu Ser Cys Ile
85 90 95

Gly Ser Pro Ile Val Ser Pro Arg Ile Val Gln Leu Glu Thr Glu Ser
100 105 110

Lys Arg Leu His Asn Lys Glu Asn Gln His Val Gln Gln Thr Leu Asn
115 120 125

Ser Thr Asn Glu Ile Glu Ala Leu Glu Thr Ser Arg Leu Tyr Glu Asp
130 135 140

Ser Gly Tyr Ser Ser Phe Ser Leu Gln Ser Gly Leu Ser Glu His Glu
145 150 155 160

Glu Gly Ser Leu Leu Glu Asn Phe Gly Asp Ser Leu Gln Ser Cys
165 170 175

Leu Leu Gln Ile Gln Ser Pro Asp Gln Tyr Pro Asn Lys Asn Leu Leu
180 185 190

Pro Val Leu His Phe Glu Lys Val Val Cys Ser Thr Leu Lys Lys Asn
195 200 205

Ala Lys Arg Asn Pro Lys Val Asp Arg Glu Met Leu Lys Glu Ile Ile
210 215 220

Ala Arg Gly Asn Phe Arg Leu Gln Asn Ile Ile Gly Arg Lys Met Gly
225 230 235 240

Leu Glu Cys Val Asp Ile Leu Ser Glu Leu Phe Arg Arg Gly Leu Arg
245 250 255

His Val Leu Ala Thr Ile Leu Ala Gln Leu Ser Asp Met Asp Leu Ile
260 265 270

Asn Val Ser Lys Val Ser Thr Thr Trp Lys Lys Ile Leu Glu Asp Asp
275 280 285

Lys Gly Ala Phe Gln Leu Tyr Ser Lys Ala Ile Gln Arg Val Thr Glu
290 295 300

Asn Asn Asn Lys Phe Ser Pro His Ala Ser Thr Arg Glu Tyr Val Met
305 310 315 320

Phe Arg Thr Pro Leu Ala Ser Val Gln Lys Ser Ala Ala Gln Thr Ser
325 330 335

Leu Lys Lys Asp Ala Gln Thr Lys Leu Ser Asn Gln Gly Asp Gln Lys
340 345 350

Gly Ser Thr Tyr Ser Arg His Asn Glu Phe Ser Glu Val Ala Lys Thr
355 360 365

Leu Lys Lys Asn Glu Ser Leu Lys Ala Cys Ile Arg Cys Asn Ser Pro
370 375 380

Ala Lys Tyr Asp Cys Tyr Leu Gln Arg Ala Thr Cys Lys Arg Glu Gly
385 390 395 400

Cys Gly Phe Asp Tyr Cys Thr Lys Cys Leu Cys Asn Tyr His Thr Thr
405 410 415

Lys Asp Cys Ser Asp Gly Lys Leu Leu Lys Ala Ser Cys Lys Ile Gly
420 425 430

Pro Leu Pro Gly Thr Lys Ser Lys Lys Asn Leu Arg Arg Leu Xaa

435

440

445

<210> 38
<211> 1050
<212> PRT
<213> Homo sapiens

<400> 38
Met Ala Ala Val Lys Lys Glu Gly Gly Ala Leu Ser Glu Ala Met Ser
5 10 15

Leu Glu Gly Asp Glu Trp Glu Leu Ser Lys Glu Asn Val Gln Pro Leu
20 25 30

Arg Gln Gly Arg Ile Met Ser Thr Leu Gln Gly Ala Leu Ala Gln Glu
35 40 45

Ser Ala Cys Asn Asn Thr Leu Gln Gln Gln Lys Arg Ala Phe Glu Tyr
50 55 60

Glu Ile Arg Phe Tyr Thr Gly Asn Asp Pro Leu Asp Val Trp Asp Arg
65 70 75 80

Tyr Ile Ser Trp Thr Glu Gln Asn Tyr Pro Gln Gly Gly Lys Glu Ser
85 90 95

Asn Met Ser Thr Leu Leu Glu Arg Ala Val Glu Ala Leu Gln Gly Glu
100 105 110

Lys Arg Tyr Tyr Ser Asp Pro Arg Phe Leu Asn Leu Trp Leu Lys Leu
115 120 125

Gly Arg Leu Cys Asn Glu Pro Leu Asp Met Tyr Ser Tyr Leu His Asn
130 135 140

Gln Gly Ile Gly Val Ser Leu Ala Gln Phe Tyr Ile Ser Trp Ala Glu
145 150 155 160

Glu Tyr Glu Ala Arg Glu Asn Phe Arg Lys Ala Asp Ala Ile Phe Gln
165 170 175

Glu Gly Ile Gln Gln Lys Ala Glu Pro Leu Glu Arg Leu Gln Ser Gln
180 185 190

His Arg Gln Phe Gln Ala Arg Val Ser Arg Gln Thr Leu Leu Ala Leu
195 200 205

Glu Lys Glu Glu Glu Glu Val Phe Glu Ser Ser Val Pro Gln Arg
210 215 220

Ser Thr Leu Ala Glu Leu Lys Ser Lys Gly Lys Lys Thr Ala Arg Ala
225 230 235 240

Pro Ile Ile Arg Val Gly Gly Ala Leu Lys Ala Pro Ser Gln Asn Arg
245 250 255

Gly Leu Gln Asn Pro Phe Pro Gln Gln Met Gln Asn Asn Ser Arg Ile

| | | | |
|---|-----|-----|-----|
| | 260 | 265 | 270 |
| Thr Val Phe Asp Glu Asn Ala Asp Glu Ala Ser Thr Ala Glu Leu Ser | | | |
| 275 | 280 | 285 | |
| Lys Pro Thr Val Gln Pro Trp Ile Ala Pro Pro Met Pro Arg Ala Lys | | | |
| 290 | 295 | 300 | |
| Glu Asn Glu Leu Gln Ala Gly Pro Trp Asn Thr Gly Arg Ser Leu Glu | | | |
| 305 | 310 | 315 | 320 |
| His Arg Pro Arg Gly Asn Thr Ala Ser Leu Ile Ala Val Pro Ala Val | | | |
| 325 | 330 | 335 | |
| Leu Pro Ser Phe Thr Pro Tyr Val Glu Glu Thr Ala Gln Gln Pro Val | | | |
| 340 | 345 | 350 | |
| Met Thr Pro Cys Lys Ile Glu Pro Ser Ile Asn His Ile Leu Ser Thr | | | |
| 355 | 360 | 365 | |
| Arg Lys Pro Gly Lys Glu Glu Gly Asp Pro Leu Gln Arg Val Gln Ser | | | |
| 370 | 375 | 380 | |
| His Gln Gln Ala Ser Glu Glu Lys Glu Lys Met Met Tyr Cys Lys | | | |
| 385 | 390 | 395 | 400 |
| Glu Lys Ile Tyr Ala Gly Val Gly Glu Phe Ser Phe Glu Glu Ile Arg | | | |
| 405 | 410 | 415 | |
| Ala Glu Val Phe Arg Lys Lys Leu Lys Glu Gln Arg Glu Ala Glu Leu | | | |
| 420 | 425 | 430 | |
| Leu Thr Ser Ala Glu Lys Arg Ala Glu Met Gln Lys Gln Ile Glu Glu | | | |
| 435 | 440 | 445 | |
| Met Glu Lys Lys Leu Lys Glu Ile Gln Thr Thr Gln Gln Glu Arg Thr | | | |
| 450 | 455 | 460 | |
| Gly Asp Gln Gln Glu Glu Thr Met Pro Thr Lys Glu Thr Thr Lys Leu | | | |
| 465 | 470 | 475 | 480 |
| Gln Ile Ala Ser Glu Ser Gln Lys Ile Pro Gly Met Thr Leu Ser Ser | | | |
| 485 | 490 | 495 | |
| Ser Val Cys Gln Val Asn Cys Cys Ala Arg Glu Thr Ser Leu Ala Glu | | | |
| 500 | 505 | 510 | |
| Asn Ile Trp Gln Glu Gln Pro His Ser Lys Gly Pro Ser Val Pro Phe | | | |
| 515 | 520 | 525 | |
| Ser Ile Phe Asp Glu Phe Leu Leu Ser Glu Lys Lys Asn Lys Ser Pro | | | |
| 530 | 535 | 540 | |
| Pro Ala Asp Pro Pro Arg Val Leu Ala Gln Arg Arg Pro Leu Ala Val | | | |
| 545 | 550 | 555 | 560 |
| Leu Lys Thr Ser Glu Ser Ile Thr Ser Asn Glu Asp Val Ser Pro Asp | | | |
| 565 | 570 | 575 | |

Val Cys Asp Glu Phe Thr Gly Ile Glu Pro Leu Ser Glu Asp Ala Ile
580 585 590

Ile Thr Gly Phe Arg Asn Val Thr Ile Cys Pro Asn Pro Glu Asp Thr
595 600 605

Cys Asp Phe Ala Arg Ala Ala Arg Phe Val Ser Thr Pro Phe His Glu
610 615 620

Ile Met Ser Leu Lys Asp Leu Pro Ser Asp Pro Glu Arg Leu Leu Pro
625 630 635 640

Glu Glu Asp Leu Asp Val Lys Thr Ser Glu Asp Gln Gln Thr Ala Cys
645 650 655

Gly Thr Ile Tyr Ser Gln Thr Leu Ser Ile Lys Lys Leu Ser Pro Ile
660 665 670

Ile Glu Asp Ser Arg Glu Ala Thr His Ser Ser Gly Phe Ser Gly Ser
675 680 685

Ser Ala Ser Val Ala Ser Thr Ser Ser Ile Lys Cys Leu Gln Ile Pro
690 695 700

Glu Lys Leu Glu Leu Thr Asn Glu Thr Ser Glu Asn Pro Thr Gln Ser
705 710 715 720

Pro Trp Cys Ser Gln Tyr Arg Arg Gln Leu Leu Lys Ser Leu Pro Glu
725 730 735

Leu Ser Ala Ser Ala Glu Leu Cys Ile Glu Asp Arg Pro Met Pro Lys
740 745 750

Leu Glu Ile Glu Lys Glu Ile Glu Leu Gly Asn Glu Asp Tyr Cys Ile
755 760 765

Lys Arg Glu Tyr Leu Ile Cys Glu Asp Tyr Lys Leu Phe Trp Val Ala
770 775 780

Pro Arg Asn Phe Ala Glu Leu Thr Val Ile Lys Val Ser Ser Gln Pro
785 790 795 800

Val Pro Trp Asp Phe Tyr Ile Asn Leu Lys Leu Lys Glu Arg Leu Asn
805 810 815

Glu Asp Phe Asp His Phe Cys Ser Cys Tyr Gln Tyr Gln Asp Gly Cys
820 825 830

Ile Val Trp His Gln Tyr Ile Asn Cys Phe Thr Leu Gln Asp Leu Leu
835 840 845

Gln His Ser Glu Tyr Ile Thr His Glu Ile Thr Val Leu Ile Ile Tyr
850 855 860

Asn Leu Leu Thr Ile Val Glu Met Leu His Lys Ala Glu Ile Val His
865 870 875 880

Gly Asp Leu Ser Pro Arg Cys Leu Ile Leu Arg Asn Arg Ile His Asp
 885 890 895
 Pro Tyr Asp Cys Asn Lys Asn Asn Gln Ala Leu Lys Ile Val Asp Phe
 900 905 910
 Ser Tyr Ser Val Asp Leu Arg Val Gln Leu Asp Val Phe Thr Leu Ser
 915 920 925
 Gly Phe Arg Thr Val Gln Ile Leu Glu Gly Gln Lys Ile Leu Ala Asn
 930 935 940
 Cys Ser Ser Pro Tyr Gln Val Asp Leu Phe Gly Ile Ala Asp Leu Ala
 945 950 955 960
 His Leu Leu Leu Phe Lys Glu His Leu Gln Val Phe Trp Asp Gly Ser
 965 970 975
 Phe Trp Lys Leu Ser Gln Asn Ile Ser Glu Leu Lys Asp Gly Glu Leu
 980 985 990
 Trp Asn Lys Phe Phe Val Arg Ile Leu Asn Ala Asn Asp Glu Ala Thr
 995 1000 1005
 Val Ser Val Leu Gly Glu Leu Ala Ala Glu Met Asn Gly Val Phe Asp
 1010 1015 1020
 Thr Thr Phe Gln Ser His Leu Asn Lys Ala Leu Trp Lys Val Gly Lys
 1025 1030 1035 1040
 Leu Thr Ser Pro Gly Ala Leu Leu Phe Gln
 1045 1050

 <210> 39
 <211> 258
 <212> PRT
 <213> Homo sapiens

 <400> 39
 Gly Lys Leu Thr Gly Ile Ser Asp Pro Val Thr Val Lys Thr Ser Gly
 5 10 15

 Ser Arg Phe Gly Ser Trp Met Thr Asp Pro Leu Ala Pro Glu Gly Asp
 20 25 30

 Asn Arg Val Trp Tyr Met Asp Gly Tyr His Asn Asn Arg Phe Val Arg
 35 40 45

 Glu Tyr Lys Ser Met Val Asp Phe Met Asn Thr Asp Asn Phe Thr Ser
 50 55 60

 His Arg Leu Pro His Pro Trp Ser Gly Thr Gly Gln Val Val Tyr Asn
 65 70 75 80

 Gly Ser Ile Tyr Phe Asn Lys Phe Gln Ser His Ile Ile Arg Phe
 85 90 95

Asp Leu Lys Thr Glu Thr Ile Leu Lys Thr Arg Ser Leu Asp Tyr Ala
 100 105 110
 Gly Tyr Asn Asn Met Tyr His Tyr Ala Trp Gly Gly His Ser Asp Ile
 115 120 125
 Asp Leu Met Val Asp Glu Ser Gly Leu Trp Ala Val Tyr Ala Thr Asn
 130 135 140
 Gln Asn Ala Gly Asn Ile Val Val Ser Arg Leu Asp Pro Val Ser Leu
 145 150 155 160
 Gln Thr Leu Gln Thr Trp Asn Thr Ser Tyr Pro Lys Arg Ser Ala Gly
 165 170 175
 Glu Ala Phe Ile Ile Cys Gly Thr Leu Tyr Val Thr Asn Gly Tyr Ser
 180 185 190
 Gly Gly Thr Lys Val His Tyr Ala Tyr Gln Thr Asn Ala Ser Thr Tyr
 195 200 205
 Glu Tyr Ile Asp Ile Pro Phe Gln Asn Lys Tyr Ser His Ile Ser Met
 210 215 220
 Leu Asp Tyr Asn Pro Lys Asp Arg Ala Leu Tyr Ala Trp Asn Asn Gly
 225 230 235 240
 His Gln Ile Leu Tyr Asn Val Thr Leu Phe His Val Ile Arg Ser Asp
 245 250 255
 Glu Leu

<210> 40
 <211> 324
 <212> PRT
 <213> Homo sapiens

<400> 40
 Met Asp Ala Pro Arg Gln Val Val Asn Phe Gly Pro Gly Pro Ala Lys
 5 10 15
 Leu Pro His Ser Val Leu Leu Glu Ile Gln Lys Glu Leu Leu Asp Tyr
 20 25 30
 Lys Gly Val Gly Ile Ser Val Leu Glu Met Ser His Arg Ser Ser Asp
 35 40 45
 Phe Ala Lys Ile Ile Asn Asn Thr Glu Asn Leu Val Arg Glu Leu Leu
 50 55 60
 Ala Val Pro Asp Asn Tyr Lys Val Ile Phe Leu Gln Gly Gly Cys
 65 70 75 80
 Gly Gln Phe Ser Ala Val Pro Leu Asn Leu Ile Gly Leu Lys Ala Gly
 85 90 95

Arg Cys Ala Asp Tyr Val Val Thr Gly Ala Trp Ser Ala Lys Ala Ala
 100 105 110
 Glu Glu Ala Lys Lys Phe Gly Thr Ile Asn Ile Val His Pro Lys Leu
 115 120 125
 Gly Ser Tyr Thr Lys Ile Pro Asp Pro Ser Thr Trp Asn Leu Asn Pro
 130 135 140
 Asp Ala Ser Tyr Val Tyr Tyr Cys Ala Asn Glu Thr Val His Gly Val
 145 150 155 160
 Glu Phe Asp Phe Ile Pro Asp Val Lys Gly Ala Val Leu Val Cys Asp
 165 170 175
 Met Ser Ser Asn Phe Leu Ser Lys Pro Val Asp Val Ser Lys Phe Gly
 180 185 190
 Val Ile Phe Ala Gly Ala Gln Lys Asn Val Gly Ser Ala Gly Val Thr
 195 200 205
 Val Val Ile Val Arg Asp Asp Leu Leu Gly Phe Ala Leu Arg Glu Cys
 210 215 220
 Pro Ser Val Leu Glu Tyr Lys Val Gln Ala Gly Asn Ser Ser Leu Tyr
 225 230 235 240
 Asn Thr Pro Pro Cys Phe Ser Ile Tyr Val Met Gly Leu Val Leu Glu
 245 250 255
 Trp Ile Lys Asn Asn Gly Ala Ala Ala Met Glu Lys Leu Ser Ser
 260 265 270
 Ile Lys Ser Gln Thr Ile Tyr Glu Ile Ile Asp Asn Ser Gln Gly Phe
 275 280 285
 Tyr Val Ser Val Gly Gly Ile Arg Ala Ser Leu Tyr Asn Ala Val Thr
 290 295 300
 Ile Glu Asp Val Gln Lys Leu Ala Ala Phe Met Lys Lys Phe Leu Glu
 305 310 315 320
 Met His Gln Leu

<210> 41
 <211> 410
 <212> PRT
 <213> Homo sapiens

<400> 41
 Met Glu Ala Glu Asn Ala Gly Ser Tyr Ser Leu Gln Gln Ala Gln Ala
 5 10 15
 Phe Tyr Thr Phe Pro Phe Gln Gln Leu Met Ala Glu Ala Pro Asn Met
 20 25 30

Ala Val Val Asn Glu Gln Gln Met Pro Glu Glu Val Pro Ala Pro Ala
 35 40 45
 Pro Ala Gln Glu Pro Val Gln Glu Ala Pro Lys Gly Arg Lys Arg Lys
 50 55 60
 Pro Arg Thr Thr Glu Pro Lys Gln Pro Val Glu Pro Lys Lys Pro Val
 65 70 75 80
 Glu Ser Lys Lys Ser Gly Lys Ser Ala Lys Pro Lys Glu Lys Gln Glu
 85 90 95
 Lys Ile Thr Asp Thr Phe Lys Val Lys Arg Lys Val Asp Arg Phe Asn
 100 105 110
 Gly Val Ser Glu Ala Glu Leu Leu Thr Lys Thr Leu Pro Asp Ile Leu
 115 120 125
 Thr Phe Asn Leu Asp Ile Val Ile Ile Gly Ile Asn Pro Gly Leu Met
 130 135 140
 Ala Ala Tyr Lys Gly His His Tyr Pro Gly Pro Gly Asn His Phe Trp
 145 150 155 160
 Lys Cys Leu Phe Met Ser Gly Leu Ser Glu Val Gln Leu Asn His Met
 165 170 175
 Asp Asp His Thr Leu Pro Gly Lys Tyr Gly Ile Gly Phe Thr Asn Met
 180 185 190
 Val Glu Arg Thr Thr Pro Gly Ser Lys Asp Leu Ser Ser Lys Glu Phe
 195 200 205
 Arg Glu Gly Gly Arg Ile Leu Val Gln Lys Leu Gln Lys Tyr Gln Pro
 210 215 220
 Arg Ile Ala Val Phe Asn Gly Lys Cys Ile Tyr Glu Ile Phe Ser Lys
 225 230 235 240
 Glu Val Phe Gly Val Lys Val Lys Asn Leu Glu Phe Gly Leu Gln Pro
 245 250 255
 His Lys Ile Pro Asp Thr Glu Thr Leu Cys Tyr Val Met Pro Ser Ser
 260 265 270
 Ser Ala Arg Cys Ala Gln Phe Pro Arg Ala Gln Asp Lys Val His Tyr
 275 280 285
 Tyr Ile Lys Leu Lys Asp Leu Arg Asp Gln Leu Lys Gly Ile Glu Arg
 290 295 300
 Asn Met Asp Val Gln Glu Val Gln Tyr Thr Phe Asp Leu Gln Leu Ala
 305 310 315 320
 Gln Glu Asp Ala Lys Lys Met Ala Val Lys Glu Glu Lys Tyr Asp Pro
 325 330 335
 Gly Tyr Glu Ala Ala Tyr Gly Gly Ala Tyr Gly Glu Asn Pro Cys Ser

| | | |
|---|-----|-----|
| 340 | 345 | 350 |
| Ser Glu Pro Cys Gly Phe Ser Ser Asn Gly Leu Ile Glu Ser Val Glu | | |
| 355 | 360 | 365 |
| Leu Arg Gly Glu Ser Ala Phe Ser Gly Ile Pro Asn Gly Gln Trp Met | | |
| 370 | 375 | 380 |
| Thr Gln Ser Phe Thr Asp Gln Ile Pro Ser Phe Ser Asn His Cys Gly | | |
| 385 | 390 | 400 |
| Thr Gln Glu Gln Glu Glu Ser His Ala | | |
| 405 | 410 | |

<210> 42
<211> 484
<212> DNA
<213> *Homo sapiens*

<210> 43
<211> 700
<212> DNA
<213> *Homo sapiens*

<210> 44
<211> 672
<212> DNA
<213> *Homo sapiens*

<210> 47

<211> 581
<212> DNA
<213> Homo sapiens

<400> 47
tcttttggaa aataaaggat ctaatgtctc cctaataagt ctctttct tccactaaaa 60
tgacctacac ggacttttat ttcttgatc aaagagggtt ttattaaggaa ctctggata 120
actatactt tactttttt taaaatggatc caaaggtaatt ttaaatgtga acaggttccc 180
ataccatgaa tgctggcttc accttctctca tcattccatc tttgaatgc aaagaaagct 240
ccctttagt ccatacttcc ttccccactc ccattcttagg atactggcc agtgcattt 300
aggcattttt tattttttt gtcggaaattt aggttattat gcttaatttg acacattaaac 360
taaatgcggc gttttaaaat atatccatca attcacgtt aataatgtctt ttgtgtcata 420
tcaaattggaa tagaatacac ttatttttt aacaatccca gaatactgtg ttttagactt 480
tgttgtgtca aaataaaatgt ttactttatc tacaaatggc aaataactgtt ttgttaaccat 540
gtatgtatgtt gtacccatca tcgaaatattat c 581

<210> 48
<211> 491
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(491)
<223> n = A,T,C or G

<400> 48
ccggggcccc cctcgagggy ttcaatggtc agatggaaaca gttgaaaggc ggggtcgaaa 60
ccctcgcatc ccacatcgcc caatctggca ttctggaaattt cgtcacaaacg atcgctcacgg 120
ccttggccaa ctttgtcgat aagctcgccg aggtcagccg gaaactctg aagtgggtca 180
cgatcatcg tgggttggcg gcggtgtctg tgccgggtc gateggcatc ggcgcgtgg 240
tctctgtcgat gggcgccctt ctccctgtca tctgtgtttt tgogagcgcc atcgccgtcg 300
tctgttcgtt catcacgcgc ggtgcattcc cagccctggc cgggtttgtt ttgtccctat 360
cgccctgtgt cgtgcccgtg gggcggttg ctgtgcagt cggccgcgtt tatctgggt 420
ggaagaactg ggcacatgtc gggccatttc tcgccaagct ttataacggc gtgaaagact 480
ggctgggtat 491

<210> 49
<211> 1929
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(1929)
<223> n = A,T,C or G

<400> 49
ttaggctagt agaggctgtt gttaatcgcc cgaggggccgc tgctagggtt gagtcggcg 60
cccggtcgcc ttggccgcgc aacaaatgtc ggcacatcggtt cgtgtggagt accgctgtcc 120
cgagtgcgcc aagggttctca gttggccgcgc caacctggcc tcgcacccqcc gctggcacaa 180
accggggccc gggccggccgc cggccggccgc gccggagccca gaagcagcag ccaggctgag 240
ggcgccgggg caccggggccgc cggcagccgc cggccacccgc cggccggggccgc 300
gagtgcgggtt cggcggaggcc gttctacggag tgccatcaact gggccaaagaa gttccgcgc 360
caggccatacc tacgcaacgca cttgtgtggc caccacccagg cgctgcagggc caaggccgc 420
ccgcgtatcgcc ccccgccgc gcacacttgc gccttgcacc cggggccgc cgagaaggcc 480

ccccaggagg cgccggcgca cgccgaggag gccggcgtgc ttgggctga gtgcgtccgs 540
 cgagtgcac ctgtgcca gtgtgcggag agtgcgtcgc cagcaaggsc gctcaggagc 600
 recrctcg cgccgtcgc cggccggccat gtgttcctct gcaagtaactg sccttgcga 660
 cctttcacat ctcggccgc cttacggcc acatcaacaat gtcacccca tccaaaaaca 720
 gagaggatgt ctctctgcg gtccggcgtgc gcccggcctg cttagcgcgc ccctccaccc 780
 cggccccccca actgtgcctt cgttggaga cccacaaaga gatgtgcgc tgcacgcccc 840
 gaaccggagt cggcgctgg ggagcctgc cccggcccc accgggtgaa agtgtcgtct 900
 ccgcttctct cgggtgtggc tgacggtaa cccataactct ctctttgtact ctcttttgaa 960
 cccccactt taacgtgtgt ctcttcggc ccccccattgc gaaacaggag taatgtctt 1020
 tctgtacaaag ggagaaaagc tgtaacgcgtt tgctctgtgg ttggaaagctt ccccttggcg 1080
 gggagaaatgt ttttttcttc ctatgtatgc ctgtgttcat gtgtctagaaa tgccgtctgg 1140
 ttcgcctcg cttaccaatc ttcgtctct atgtatgtatgc ctgtacgggtt gttttttgtt 1200
 aatcttgagg aataaatgc ttatatttc acaggctgtat aattgaactt cccacacgt 1260
 tagctttattt atggctgtgt aactgtgttgc gtctggctt acctttttgtt atgtgaacaa 1320
 atcaaatgtt tttttttttt tgatagccac aatgtcttg aactgttgtt 1380
 tgggatgtt ttgtgggggg agggaaggga gtgttccgaa gatgtgtatg taactgcctc 1440
 agtgtttcac gtaagactt ttgtttttt catctttttt gaggtaggtt tatttttttttcc 1500
 ctcttaatgtt atatgtgtt ttatgtatgtt ttgttattttt ttctttttt attttttttt 1560
 attatgtatgtt ttatgtatgtt atttgtatgtt agatgtatgtt ggccgcgtcgc cccctcccca 1620
 ccctgcacat cttagacatc cactgtgcgtt tttagaaagag agccctttt taaagggtt 1680
 tgcttaaagt tttaactttt atacctatot gatgtatgtt cagacaacctt atattttttt 1740
 ctgtctcgag ggtcccccagg gcccctgtac aaccgcacgc tctttactttt aaatgcatac 1800
 tctttttctac atacattttt ttcttaattt ttagatgtt atagaaagct tcaatagaac 1860
 tgtttcaact gtataacttat ttaactattca aataaaaat tttcaagtc aaaaaaaaaa 1920
 aaaaaaaaaaag 1929

<210> 50
 <211> 6183
 <212> DNA
 <213> Homo sapiens

<400> 50
 ctttttgtat ggagaaggc aggatgtttt taactgaatg tgacccagg ggaataactag 60
 agaaaaataat aaaatttcgt aatggggcag cgtggagaaa tcctaaagaga aatagcataa 120
 gagcattttt gcacacatcc agggaaatgt aactttcgc acacccgtatg acgttgcgc 180
 gttaaaggag tgatgtatgc ttccatgttc agatccgtatg gcttttttaggg aaggaactac 240
 agttgtgcac tttaatgtatgtt gcacatgtatc tttaatgttgc tgtaaaatcc tacaagaaac 300
 acagaaatctt atgtatgtatgtt agtgcgtatc ttgtatgtatc tgcaaaagggtt agagaaggac 360
 ggacggtaga aattttctgc aagaaatgtt gatgtatgttgc ttatgtatctt aataaagaca 420
 aagtcatat tttatgtatgttgc tcgacacagc agggcaatgtt gatgtatgttgc ttttttttttcc 480
 tggggaaaaa gagacgttgc ctttttttttcc ttttttttttcc ttttttttttcc ttttttttttcc 540
 gtgcgtctgg gcgatgtgttgc agagacgttgc ttatgtatgttgc tgatgtatgttgc ttttttttttcc 600
 aaagactcgtt atgtatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 660
 caggacttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 720
 cgttctaaatg gaaacggatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 780
 gatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 840
 aaaaaggaaat ccctgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 900
 gaccgcgttgc acaggacatgc gacatgtatgttgc ttatgtatgttgc ttatgtatgttgc 960
 gagggaaatctt tgatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 1020
 agtaggtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 1080
 ttggggaaatctt tgatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 1140
 ggcattcgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 1200
 gatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 1260
 gaaagacttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 1320
 accaaacacatc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 1380
 gttctgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 1440
 gaaacagatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc ttatgtatgttgc 1500

ccctcattcc ctgacgttga ggaaaaagat agcgagagcc tggcagtaat gacccaagg 1560
 ggttagtcgg tggaaaaggaa caaaggaaat tttaatggttc tgaggcaggc atactgcgtc 1620
 cagctggcgg gagggttgtt ttccataaa acctacaagg aaacctggat gtatccgtgc 1680
 gacccatctg caggggaaag gggcaaacc aaagttagtcg acatgggtgg aacacaatac 1740
 tttaaataata aaccatccaccc aaggctgtaa gagggggaa ccaatggggc gatccctggc 1800
 tttgtatgcg cggcggacgt gacagggtcc tacccggcacc acccggccgt tcgggggtc 1860
 cccccaaagg tgcgggttgc ctggaaattt ctggccatgc atgaaaatgt gtcacatgt 1920
 cccacccggg gatgcacagg aagggtgtt ctggacacca acggccacac ccacagggt 1980
 ctttctgtt gttcaatgc tgcaatgtt gaaaatggcaaa tgccgggggg taaaataatgc 2040
 ctgttgcctt cccaaatactgg cagttgtcc acggggccc acggacacaa ttgttgatgt 2100
 caaatgtat ttcaatgttgc gtcacacggc atccatccatc ccacggccgt atgttgcataa 2160
 gaacaagaga aqttttggaa agtaccatcc gtttatgtcc gttttgtatc coaagttttc 2220
 gttttttttttt ctctcatata aacagtgcggaa gggacaaaaaa caccacccatc ttgttgcataa 2280
 aatggatccat ttccaaatccatc gaaaattttttcc atactgcgtcc ctgtatgcgg ccggccaccc 2340
 cagagccctg ggcgttgcgg ctcttatacgc taagggtcaat tgatgttgcg caccacacata 2400
 gcaacggctc ctggccatccctt gaaacccatcc acggcgatgc gggaaaggcc acacatccctc 2460
 ttccaaacaaagc cagacatgtt gcatggccaa gggacccaaa tagaaatggta tgaaatgttc 2520
 attttttttttttaaactggat aaaaaatgtca atccgtggaca atgttgcgttccatccatc 2580
 tctaaatccat ttatccatccatc ttccatccatc aaaaacaaagc atttttgttc 2640
 aatggcggat tttatccatccatc ttccatccatc aaaaacaaagc atttttgttc 2700
 aacggaaatcc acggggaaatc agaggaggaa agaaagaaaaa accggatgtt cccttgcgtat 2760
 aatttttaggg aaaaaaaatgttccatccatc ttccatccatc aaaaacaaagc atttttgttc 2820
 catgcaagag atctttttttt gaaactaatac acctgttccaa caccaggatg tgatggaaatg 2880
 gggccatgtt cggggaaacta tgcatgttccatc gggatgtttt ccgtatgttccatccatc 2940
 aagactctaa aatccctccat ggttgcggccaa ttccatccatc aaaaacaaagc atttttgttc 3000
 tgccatgttgc cggggccatgtt gactggaaaat tatgttccatc aaaaacaaagc atttttgttc 3060
 ctccgttgcggaa gggaaatgttgc ttccatccatc aaaaacaaagc atttttgttc 3120
 gaaacttgcataatc ttccatccatc aaaaacaaagc atttttgttc 3180
 aatttttaggg aaaaaaaatgttccatccatc ttccatccatc aaaaacaaagc atttttgttc 3240
 atggatccatc cccttccatccatc gaaacttgcataatc ttccatccatc aaaaacaaagc atttttgttc 3300
 ggttgcgtatg ggttgcggccaa ttccatccatc aaaaacaaagc atttttgttc 3360
 tgatgttgcataatc ttccatccatc aaaaacaaagc atttttgttc 3420
 aagtttttttttccatccatc aaaaacaaagc atttttgttc 3480
 ataaggaaatcc tttatccatccatc aaaaacaaagc atttttgttc 3540
 cagatccatccatc ttccatccatc aaaaacaaagc atttttgttc 3600
 caacaaatgtt gaaatgttgc gggatgtttt ccacccatccatc aaaaacaaagc atttttgttc 3660
 ctgttgcgtatc ttccatccatc aaaaacaaagc atttttgttc 3720
 gtaatccatccatc ttccatccatc aaaaacaaagc atttttgttc 3780
 gtttttttttttccatccatc aaaaacaaagc atttttgttc 3840
 gtttttttttttccatccatc aaaaacaaagc atttttgttc 3900
 ttttttttttttccatccatc aaaaacaaagc atttttgttc 3960
 tgatgttgcataatc ttccatccatc aaaaacaaagc atttttgttc 4020
 tgatgttgcataatc ttccatccatc aaaaacaaagc atttttgttc 4080
 tgatgttgcataatc ttccatccatc aaaaacaaagc atttttgttc 4140
 tactaatccatc aagtttttttccatccatc aaaaacaaagc atttttgttc 4200
 gtttttttttttccatccatc aaaaacaaagc atttttgttc 4260
 ttttttttttttccatccatc aaaaacaaagc atttttgttc 4320
 ttttttttttttccatccatc aaaaacaaagc atttttgttc 4380
 tgactatccatc ttccatccatc aaaaacaaagc atttttgttc 4440
 gggccatgttccatccatc aaaaacaaagc atttttgttc 4500
 cggggccatgttccatccatc aaaaacaaagc atttttgttc 4560
 cggggccatgttccatccatc aaaaacaaagc atttttgttc 4620
 ctttttttttttccatccatc aaaaacaaagc atttttgttc 4680
 tgatgttgcataatc ttccatccatc aaaaacaaagc atttttgttc 4740
 ttttttttttttccatccatc aaaaacaaagc atttttgttc 4800
 aatggaaaatccatccatc aaaaacaaagc atttttgttc 4860
 atggaaaatccatccatc aaaaacaaagc atttttgttc 4920
 atggaaaatccatccatc aaaaacaaagc atttttgttc 4980

```
<210> 51
<211> 1704
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(1704)
<223> n = A.T.C or G
```

```

cattattaaa aaaaaattc tggatgaaa actgtctatga taaagtgcgttgcgttggatgtg 1500
gggtttttgtatcagac agacggaaaa ttccgggttc ttccgggttggatgttggatgtcgt 1560
ttgtacatggaaatcgta ggcggaaatggatggatggatggatggatggatggatggatggatgg 1620
agaatgttggaaatggatggatggatggatggatggatggatggatggatggatggatggatggatgg 1680
aaacatgttttctttagac ctat                                         1704

```

<210> 52
<211> 1886
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<222> (1)...(1886)
<223> n = A,T,C or G

<210> 53
<211> 877
<212> DNA
<213> *Homo sapiens*

<220>

<221> misc_feature
<222> (1)...(877)
<223> n = A,T,C or G

<210> 54
<211> 1364
<212> DNA
<213> Homo sapiens

```
<220>
<221> misc_feature
<222> (1)...(1364)
<223> n = A,T,C or G
```

<210> 55
<211> 539
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(539)
<223> n = A,T,C or G

<400> 55
ccggcccccc cctcgagggy ttcaatggc agatgaaaca gttgaaaggc gccgtcgaaa 60
ccctcgccat ccaatcgccg caaatctggca ttctggaaatt cgtcacaaacg atcgtcaccc 120
ccttggccaa ctttgtcgat aagctcgccg aggtcagccc ggaaactctg aagtgggtca 180
cgatcatcg tgggtggccg gccgtgtcgat gtccgggtgc gatcgccatc ggcgcccgtgg 240
tcctcgctgtt gggcgccatc ctccctgtca tcgtgcgtgt tcgcagccgc atccggcgctg 300
tcgttcgtt catcacggcc ggtgcattcc cagccctggc cgggttgcgtt gttccctat 360
cgccctgtct cgtccgcgtg gccgcgggtc ctgtcgatc cgccgcgtt tatctgtgt 420
ggaaagactg ggacatgtc gggccattc tcgcccaagct ttataaacgga gtaaagacgt 480
ggctggtcga taagctcgcc aagggtgtggg aaactctcaa gagcaagata aaagccgtaa 539

<210> 56
<211> 510
<212> PRT
<213> Homo sapiens

<400> 56
Met Pro Arg Gly Phe Leu Val Lys Arg Ser Lys Lys Ser Thr Pro Val
5 10 15

Ser Tyr Arg Val Arg Gly Gly Glu Asp Gly Asp Arg Ala Leu Leu Leu
20 25 30

Ser Pro Ser Cys Gly Gly Ala Arg Ala Glu Pro Pro Ala Pro Ser Pro
35 40 45

Val Pro Gly Pro Leu Pro Pro Pro Pro Ala Glu Arg Ala His Ala
50 55 60

Ala Leu Ala Ala Ala Leu Ala Cys Ala Pro Gly Pro Gln Pro Pro Pro
65 70 75 80

Gln Gly Pro Arg Ala Ala His Phe Gly Asn Pro Glu Ala Ala His Pro
85 90 95

Ala Pro Leu Tyr Ser Pro Thr Arg Pro Val Ser Arg Glu His Glu Lys
100 105 110

His Lys Tyr Phe Glu Arg Ser Phe Asn Leu Gly Ser Pro Val Ser Ala
115 120 125

Glu Ser Phe Pro Thr Pro Ala Ala Leu Leu Gly Gly Gly Gly Gly
130 135 140

Gly Ala Ser Gly Ala Gly Gly Gly Thr Cys Gly Gly Asp Pro Leu
145 150 155 160

Leu Phe Ala Pro Ala Glu Leu Lys Met Gly Thr Ala Phe Ser Ala Gly
 165 170 175
 Ala Glu Ala Ala Arg Gly Pro Gly Pro Pro Leu Pro Pro Ala
 180 185 190
 Ala Ala Leu Arg Pro Pro Gly Lys Arg Pro Pro Pro Thr Ala Ala
 195 200 205
 Glu Pro Pro Ala Lys Ala Val Lys Ala Pro Gly Ala Lys Lys Pro Lys
 210 215 220
 Ala Ile Arg Lys Leu His Phe Glu Asp Glu Val Thr Thr Ser Pro Val
 225 230 235 240
 Leu Gly Leu Lys Ile Lys Glu Gly Pro Val Glu Ala Pro Arg Gly Arg
 245 250 255
 Ala Gly Gly Ala Ala Arg Pro Leu Gly Glu Phe Ile Cys Gln Leu Cys
 260 265 270
 Lys Glu Glu Tyr Ala Asp Pro Phe Ala Leu Ala Gln His Lys Cys Ser
 275 280 285
 Arg Ile Val Arg Val Glu Tyr Arg Cys Pro Glu Cys Ala Lys Val Phe
 290 295 300
 Ser Cys Pro Ala Asn Leu Ala Ser His Arg Arg Trp His Lys Pro Arg
 305 310 315 320
 Pro Ala Pro Ala Ala Arg Ala Pro Glu Pro Glu Ala Ala Ala Arg
 325 330 335
 Ala Glu Ala Arg Glu Ala Pro Gly Gly Ser Asp Arg Asp Thr Pro
 340 345 350
 Ser Pro Gly Gly Val Ser Glu Ser Gly Ser Glu Asp Gly Leu Tyr Glu
 355 360 365
 Cys His His Cys Ala Lys Lys Phe Arg Arg Gln Ala Tyr Leu Arg Lys
 370 375 380
 His Leu Leu Ala His His Gln Ala Leu Gln Ala Lys Gly Ala Pro Leu
 385 390 395 400
 Ala Pro Pro Ala Glu Asp Leu Leu Ala Leu Tyr Pro Gly Pro Asp Glu
 405 410 415
 Lys Ala Pro Gln Glu Ala Ala Gly Asp Gly Glu Gly Ala Gly Val Leu
 420 425 430
 Gly Leu Ser Ala Ser Ala Glu Cys His Leu Cys Pro Val Cys Gly Glu
 435 440 445
 Ser Phe Ala Ser Lys Gly Ala Gln Glu Arg His Leu Arg Leu Leu His
 450 455 460
 Ala Ala Gln Val Phe Pro Cys Lys Tyr Cys Pro Ala Thr Phe Tyr Ser

| | | | |
|---|-----|-----|-----|
| 465 | 470 | 475 | 480 |
| Ser Pro Gly Leu Thr Arg His Ile Asn Lys Cys His Pro Ser Glu Asn | | | |
| 485 | 490 | 495 | |
| Arg Gln Val Ile Leu Leu Gln Val Pro Val Arg Pro Ala Cys | | | |
| 500 | 505 | 510 | |
| <210> 57 | | | |
| <211> 1047 | | | |
| <212> PRT | | | |
| <213> Homo sapiens | | | |
| <400> 57 | | | |
| Met Asp Ala Glu Ala Glu Asp Lys Thr Leu Arg Thr Arg Ser Lys Gly | | | |
| 5 | 10 | 15 | |
| Thr Glu Val Pro Met Asp Ser Leu Ile Gln Glu Leu Ser Val Ala Tyr | | | |
| 20 | 25 | 30 | |
| Asp Cys Ser Met Ala Lys Lys Arg Thr Ala Glu Asp Gln Ala Leu Gly | | | |
| 35 | 40 | 45 | |
| Val Pro Val Asn Lys Arg Lys Ser Leu Leu Met Lys Pro Arg His Tyr | | | |
| 50 | 55 | 60 | |
| Ser Pro Lys Ala Asp Cys Gln Glu Asp Arg Ser Asp Arg Thr Glu Asp | | | |
| 65 | 70 | 75 | 80 |
| Asp Gly Pro Leu Glu Thr His Gly His Ser Thr Ala Glu Glu Ile Met | | | |
| 85 | 90 | 95 | |
| Ile Lys Pro Met Asp Glu Ser Leu Leu Ser Thr Ala Gln Glu Asn Ser | | | |
| 100 | 105 | 110 | |
| Ser Arg Lys Glu Asp Arg Tyr Ser Cys Tyr Gln Glu Leu Met Val Lys | | | |
| 115 | 120 | 125 | |
| Ser Leu Met His Leu Gly Lys Phe Glu Lys Asn Val Ser Val Gln Thr | | | |
| 130 | 135 | 140 | |
| Val Ser Glu Asn Leu Asn Asp Ser Gly Ile Gln Ser Leu Lys Ala Glu | | | |
| 145 | 150 | 155 | 160 |
| Ser Asp Glu Ala Asp Glu Cys Phe Leu Ile His Ser Asp Asp Gly Arg | | | |
| 165 | 170 | 175 | |
| Asp Lys Ile Asp Asp Ser Gln Pro Pro Phe Cys Ser Ser Asp Asp Asn | | | |
| 180 | 185 | 190 | |
| Glu Ser Asn Ser Glu Ser Ala Glu Asn Gly Trp Asp Ser Gly Ser Asn | | | |
| 195 | 200 | 205 | |
| Phe Ser Glu Glu Thr Lys Pro Pro Arg Val Pro Lys Tyr Val Leu Thr | | | |
| 210 | 215 | 220 | |
| Asp His Lys Lys Asp Leu Leu Glu Val Pro Glu Ile Lys Thr Glu Gly | | | |

| | | | |
|---|-----|-----|-----|
| 225 | 230 | 235 | 240 |
| Asp Lys Phe Ile Pro Cys Glu Asn Arg Cys Asp Ser Glu Thr Glu Arg | | | |
| 245 | 250 | 255 | |
| Lys Asp Pro Gln Asn Ala Leu Ala Glu Pro Leu Asp Gly Asn Ala Gln | | | |
| 260 | 265 | 270 | |
| Pro Ser Phe Pro Asp Val Glu Glu Glu Asp Ser Glu Ser Leu Ala Val | | | |
| 275 | 280 | 285 | |
| Met Thr Glu Glu Gly Ser Asp Leu Glu Lys Ala Lys Gly Asn Leu Ser | | | |
| 290 | 295 | 300 | |
| Leu Leu Glu Gln Ala Ile Ala Leu Gln Ala Glu Arg Gly Cys Val Phe | | | |
| 305 | 310 | 315 | 320 |
| His Asn Thr Tyr Lys Glu Leu Asp Arg Phe Leu Leu Glu His Leu Ala | | | |
| 325 | 330 | 335 | |
| Gly Glu Arg Arg Gln Thr Lys Val Ile Asp Met Gly Arg Gln Ile | | | |
| 340 | 345 | 350 | |
| Phe Asn Asn Lys His Ser Pro Arg Pro Glu Lys Arg Glu Thr Lys Cys | | | |
| 355 | 360 | 365 | |
| Pro Ile Pro Gly Cys Asp Gly Thr Gly His Val Thr Gly Leu Tyr Pro | | | |
| 370 | 375 | 380 | |
| His His Arg Ser Leu Ser Gly Cys Pro His Lys Val Arg Val Pro Leu | | | |
| 385 | 390 | 395 | 400 |
| Glu Ile Leu Ala Met His Glu Asn Val Leu Lys Cys Pro Thr Pro Gly | | | |
| 405 | 410 | 415 | |
| Cys Thr Gly Arg Gly His Val Asn Ser Asn Arg Asn Thr His Arg Ser | | | |
| 420 | 425 | 430 | |
| Leu Ser Gly Cys Pro Ile Ala Ala Ala Glu Lys Leu Ala Met Ser Gln | | | |
| 435 | 440 | 445 | |
| Asp Lys Asn Gln Leu Asp Ser Pro Gln Thr Gly Gln Cys Pro Asp Gln | | | |
| 450 | 455 | 460 | |
| Ala His Arg Thr Ser Leu Val Lys Gln Ile Glu Phe Asn Phe Pro Ser | | | |
| 465 | 470 | 475 | 480 |
| Gln Ala Ile Thr Ser Pro Arg Ala Thr Val Ser Lys Glu Gln Glu Lys | | | |
| 485 | 490 | 495 | |
| Phe Gly Lys Val Pro Phe Asp Tyr Ala Ser Phe Asp Ala Gln Val Phe | | | |
| 500 | 505 | 510 | |
| Gly Lys Arg Pro Leu Ile Gln Thr Val Gln Gly Arg Lys Thr Pro Pro | | | |
| 515 | 520 | 525 | |
| Phe Pro Glu Ser Lys His Phe Pro Asn Pro Val Lys Phe Pro Asn Arg | | | |
| 530 | 535 | 540 | |

Leu Pro Ser Ala Gly Ala His Thr Gln Ser Pro Gly Arg Ala Ser Ser
545 550 555 560

Tyr Ser Tyr Gly Gln Cys Ser Glu Asp Thr His Ile Ala Ala Ala Ala
565 570 575

Ala Ile Leu Asn Leu Ser Thr Arg Cys Arg Glu Ala Thr Asp Ile Leu
580 585 590

Ser Asn Lys Pro Gln Ser Leu His Ala Lys Gly Ala Glu Ile Glu Val
595 600 605

Asp Glu Asn Gly Thr Leu Asp Leu Ser Met Lys Lys Asn Arg Ile Leu
610 615 620

Asp Lys Ser Ala Pro Leu Thr Ser Ser Asn Thr Ser Ile Pro Thr Pro
625 630 635 640

Ser Ser Ser Pro Phe Lys Thr Ser Ser Ile Leu Val Asn Ala Ala Phe
645 650 655

Tyr Gln Ala Leu Cys Asp Gln Glu Gly Trp Asp Thr Pro Ile Asn Tyr
660 665 670

Ser Lys Thr His Gly Lys Thr Glu Glu Glu Lys Glu Lys Asp Pro Val
675 680 685

Ser Ser Leu Glu Asn Leu Glu Glu Lys Lys Phe Pro Gly Glu Ala Ser
690 695 700

Ile Pro Ser Pro Lys Pro Lys Leu His Ala Arg Asp Leu Lys Lys Glu
705 710 715 720

Leu Ile Thr Cys Pro Thr Pro Gly Cys Asp Gly Ser Gly His Val Thr
725 730 735

Gly Asn Tyr Ala Ser His Arg Ser Val Ser Gly Cys Pro Leu Ala Asp
740 745 750

Lys Thr Leu Lys Ser Leu Met Ala Ala Asn Ser Gln Glu Leu Lys Cys
755 760 765

Pro Thr Pro Gly Cys Asp Gly Ser Gly His Val Thr Gly Asn Tyr Ala
770 775 780

Ser His Arg Ser Leu Ser Gly Cys Pro Arg Ala Arg Lys Gly Gly Val
785 790 795 800

Lys Met Thr Pro Thr Lys Glu Glu Lys Glu Asp Pro Glu Leu Lys Cys
805 810 815

Pro Val Ile Gly Cys Asp Gly Gln Gly His Ile Ser Gly Lys Tyr Thr
820 825 830

Ser His Arg Thr Ala Ser Gly Cys Pro Leu Ala Ala Lys Arg Gln Lys
835 840 845

Glu Asn Pro Leu Asn Gly Ala Ser Leu Ser Trp Lys Leu Asn Lys Gln
 850 855 860
 Glu Leu Pro His Cys Pro Leu Pro Gly Cys Asn Gly Leu Gly His Val
 865 870 875 880
 Asn Asn Val Phe Val Thr His Arg Ser Leu Ser Gly Cys Pro Leu Asn
 885 890 895
 Ala Gln Val Ile Lys Lys Gly Lys Val Ser Glu Glu Leu Met Thr Ile
 900 905 910
 Lys Leu Lys Ala Thr Gly Gly Ile Glu Ser Asp Glu Glu Ile Arg His
 915 920 925
 Leu Asp Glu Glu Ile Lys Glu Leu Asn Glu Ser Asn Leu Lys Ile Glu
 930 935 940
 Ala Asp Met Met Lys Leu Gln Thr Gln Ile Thr Ser Met Glu Ser Asn
 945 950 955 960
 Leu Lys Thr Ile Glu Glu Glu Asn Lys Leu Ile Glu Gln Asn Asn Glu
 965 970 975
 Ser Leu Leu Lys Glu Leu Ala Gly Leu Ser Gln Ala Leu Ile Ser Ser
 980 985 990
 Leu Ala Asp Ile Gln Leu Pro Gln Met Gly Pro Ile Ser Glu Gln Asn
 995 1000 1005
 Phe Glu Ala Tyr Val Asn Thr Leu Thr Asp Met Tyr Ser Asn Leu Glu
 1010 1015 1020
 Arg Asp Tyr Ser Pro Glu Cys Lys Ala Leu Leu Glu Ser Ile Lys Gln
 1025 1030 1035 1040
 Ala Val Lys Gly Ile His Val
 1045

<210> 58
 <211> 2165
 <212> DNA
 <213> Homo sapiens

<400> 58
 cgccacgcgt ggggtcgccg aggccggcgc gatcgccgag ctgtgccgg gccgcgtgtc 60
 gggcatctcg gtggccatcg cgcacgggtt ctctcgccg tccctcaaca tcttgctcaa 120
 gttccctatc agccgcatac agtttcctct cctgaccctt gttcagtgtc tgaccagetc 180
 cacgggggg ctgagcttgg agctgtctcg ggcgcctggg ctatcgccg tgccccctt 240
 cggctcgacg ctggcgccgt ctttcgcggg ggtcgccgtg ctctccacccg tgcaagtccag 300
 cttcaacgtc ttgtccctgc gggccctcg cttgcggatc tacgttgtct tcaaggcgctg 360
 cttgcacctg gtaccatgc tcattcgccgt cttgtgtctc aagaacggcg cgccctcgcc 420
 aggggtctg ggtcgccgtc tcattcaccc ctgcggccgc ggcctggcag gagccggccga 480
 cttgcgggc gacccatcg ggtacgttc gggagtgtc ggggtgttgc tgacgcgtc 540
 ctatcggtt ctatccaga agggccatcg agacaccggg cacggccgc tcacccgcga 600
 gtacgtatc ggcgtctcg ccaccccgct gctgttcattc tgctccctcg ccacccggaa 660
 ctccatccac gcttggactc tccgggtctg gaaggaccgc gccatggtct gcatttcgt 720

ggcctgcata ctgatcgct ggcctatgaa cttcaccacg ctgcaactgca cctacatcaa 780
 ttccggctg accacccttc tggtcatgac cggcgtgtt gtgaacaccc tggctcttat 840
 cattactgt gtggccaagt tcatggagac cagaaggaa agcaactacg aggacctgaa 900
 ggcccaccc tggggagagg aggcgcagct aagtggagac cagctgccgt tcgtgatgg 960
 ggagctgccccc gggggaggag gaatggccg gtccggaaatgtt gggggaggccg cagggtggccc 1020
 cgctcaggag agcaggcaag aggtcagggtt cagccccca ggagtccccc tgggtggctt 1080
 gagctctgaa gaaggggaca ggagggtttt aaaaatgtct tacctcgagg tatggagggtt 1140
 gtttagggaa accaggtaa tgaaaggaa ttatgttgc gaaaacgggg agtttacccca 1200
 tccttgagaa ggagggtcat gtacgtact atgtcgatcat actttttatataatgatgaa 1260
 atgacgttgtt ttaatggag gtcctcccggtt ttatgttgc ggggggtttt ggggggggg 1320
 aaagaaaggaa gctgaaagggtt actgcacacgg agcaacaaaa tttagcacctgttggatatt 1380
 ttatgtgtac ttccatgttgc gatcacaagg gacaaaaaaa ttgtggatcta cttaaacaag 1440
 taaggcaacg ttttttttttcc agactctgttcc caccatctt ttttttttttcc attaaccat 1500
 ctaaatatct gcatgttccca aggttccca acggccccc tccaaatggat gatgttagaa 1560
 atgatgacacg cacttagtaa gttcaaaatgtt gacattcagg gatgtttt ttgtatgatag 1620
 aactacatgtt tttatgttgc gtcggccaa ggtttttttt gttttttttt gttttttttt 1680
 attgtatgtt gttttttttt tttatgtatgtt gttttttttt tttttttttt tttttttttt 1740
 cacttggttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1800
 agctttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1860
 aattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1920
 ctgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1980
 ttccggatata actgaaaccctt agccggaa gggaccactt gtaaaggaaatggcc tttttttttt 2040
 ttccactgtca catcccaac catgttacaa aagagacac tttttttttt tttttttttt 2100
 attttttttttctt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2160
 aaaaaa 2165

<210> 59

<211> 1176

<212> DNA

<213> Homo sapiens

<400> 59

atggggaaatc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 60
 ttctcgggtt ccctcaacat cttgtctcaat ttccatcatca gcccgttccca gttttttttt 120
 ctgacccctgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 180
 cggccggccgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 240
 gtccgggttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 300
 ctggccatgt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 360
 ctgtgttccca agaaaggccgtt gggccgttccca ggggttccca gggccgttccca 420
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 480
 gggtgtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 540
 gacccggccgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 600
 ctgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 660
 aaggacccggccgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 720
 ttccaccaatgt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 780
 ggcgtgtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 840
 agaaaggccgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 900
 agtggggaccc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 960
 tcagaagggtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1020
 agcccccggccgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1080
 aaagatgtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1140
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1176
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt

<210> 60

<211> 1089

<212> DNA

<213> Homo sapiens

<400> 60
 ggcacatcg ggtgcggcgc aggcggccgc gatgcggcag ctgtgcggg gccgcgtgct 60
 ggcacatcg gtggccatcg cgcacgggtt ttatcggtc tccctaaaca tttgtctaa 120
 gttcctatc agccgtacc agttctccct cttgaccctg gtgcagtgc tgaccgc 180
 cacccggcgc ctgagccatgg agtgcgtgcg ggcgcctcggt ctcatcgccg tgccccctt 240
 cggtctgagc ctggccgcgtt ctttcgcggg ggtgcgggtt ctctccacgc tgcatcgac 300
 cctcaecgtc tggcccttcgc gggccatcg cttgcggatg tacgtggct tcaagcgctg 360
 cctgccttcgtc gtccatcgatc cttatcggtc aagaacggcg ccgcctcgcc 420
 agggatgtcgccggccgtc tcatcacccatc ctgcggccgc gccttgccg gagccggcga 480
 ctgtgcggc gacccatcg gtatcgatc gggatgtctg ggcgtctgg tgacgcgtc 540
 ctatccatgt ctatccatgg agggccacggc agacaccggc cacggggccgc tcaccgcgca 600
 gtacgtatcgc ggcgtctgc ccacccatgt ctgtgtatc tgctcttcgc ccacccatgc 660
 ctccatccac gcttgcactt tcccccgtt gaaggacccgc gccatgttgtt gcatcttcgt 720
 ggatgtatcgc ctatcgatgtt ggcgcgtatc cttcacccatgc ctgcactgtca cttatcatcaa 780
 ttccggccgtt accacatcttc tttatcgatc cggcgtgtgtt gtgaacaccc tgggtctaa 840
 cattatctgtt gggccaaatc tttatcgatgtt cagaagccaa agcaactacg aggacatgg 900
 ggccacgtt cggggaggagg agggccacgtt aagtggagac cagtcgttgcgt tggtatgg 960
 ggatgtatcgc gggggggggg gaaatggccgtt gtcaagatgtt gggggggggc cagggtggcc 1020
 ctgcgtatcgc agcaggcaag aggttcggggc cagccccccgc ggagtcggcc tgggtggctg 1080
 gagtcgtatc 1089

<210> 61

<211> 362

<212> PRT

<213> Homo sapiens

<400> 61

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg | His | Arg | Trp | Val | Arg | Arg | Gly | Arg | Arg | Asp | Ala | Ala | Ala | Val | Pro |
| | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | 10 | | | | 15 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gly | Pro | Arg | Ala | Gly | His | Leu | Gly | Gly | His | Arg | Ala | Arg | Gly | Leu | Leu |
| | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | 25 | | | | 30 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gly | Leu | Pro | Gln | His | Leu | Ala | Gln | Val | Pro | His | Gln | Pro | Leu | Pro | Val |
| | | | | | | | | | | | | | | | |
| 35 | | | | | | | | | | | 40 | | | | 45 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Leu | Leu | Pro | Asp | Pro | Gly | Ala | Val | Pro | Asp | Gln | Leu | His | Arg | Gly | Ala |
| | | | | | | | | | | | | | | | |
| 50 | | | | | | | | | | | 55 | | | | 60 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Glu | Pro | Gly | Ala | Ala | Ala | Ala | Pro | Arg | Ala | His | Arg | Arg | Ala | Pro | Leu |
| | | | | | | | | | | | | | | | |
| 65 | | | | | | | | | | | 70 | | | | 80 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg | Ser | Glu | Pro | Gly | Ala | Leu | Leu | Arg | Gly | Gly | Arg | Gly | Ala | Leu | His |
| | | | | | | | | | | | | | | | |
| 85 | | | | | | | | | | | 90 | | | | 95 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ala | Ala | Val | Gln | Pro | His | Ala | Leu | Val | Pro | Ala | Arg | Pro | Gln | Pro | Ala |
| | | | | | | | | | | | | | | | |
| 100 | | | | | | | | | | | 105 | | | | 110 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| His | Val | Arg | Gly | Leu | Gln | Ala | Leu | Pro | Ala | Pro | Gly | His | His | Ala | His |
| | | | | | | | | | | | | | | | |
| 115 | | | | | | | | | | | 120 | | | | 125 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg | Arg | Pro | Gly | Ala | Gln | Glu | Arg | Arg | Ala | Leu | Ala | Arg | Gly | Ala | Gly |
| | | | | | | | | | | | | | | | |
| 130 | | | | | | | | | | | 135 | | | | 140 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gly | Gly | Ala | His | His | His | Leu | Arg | Arg | Arg | Pro | Gly | Arg | Ser | Arg | Arg |
| | | | | | | | | | | | | | | | |
| 145 | | | | | | | | | | | 150 | | | | 155 |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Pro | Asp | Gly | Arg | Pro | His | Arg | Val | Arg | His | Gly | Ser | Ala | Gly | Gly | Ala |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

| 165 | 170 | 175 |
|--|-----|-----|
| Gly Ala Arg Cys Leu Pro Gly Ala His Pro Glu Gly Gln Arg Arg His 180 | 185 | 190 |
| Arg Ala Arg Ala Ala His Arg Ala Val Arg His Arg Arg Leu Cys His 195 | 200 | 205 |
| Pro Ala Ala Gly His Leu Leu Leu Arg Gln His Arg Leu His Pro Arg 210 | 215 | 220 |
| Leu Asp Leu Pro Gly Leu Glu Gly Pro Gly His Gly Leu His Leu Arg 225 | 230 | 235 |
| Gly Leu His Pro Asp Arg Leu Arg His Glu Leu His His Ala Ala Leu 245 | 250 | 255 |
| His Leu His Gln Phe Gly Arg Asp His Leu Ser Val His Cys Arg Arg 260 | 265 | 270 |
| Gly Gly Glu His Pro Gly Leu Tyr His Leu Leu Cys Gly Gln Val His 275 | 280 | 285 |
| Gly Asp Gln Lys Ala Lys Gln Leu Arg Gly Pro Gly Gly Pro Ala Ser 290 | 295 | 300 |
| Gly Arg Gly Gly Ala Ala Lys Trp Arg Pro Ala Ala Val Arg Asp Gly 305 | 310 | 315 |
| Gly Ala Ala Arg Gly Gly Arg Lys Trp Pro Val Arg Arg Trp Gly Gly 325 | 330 | 335 |
| Ser Arg Trp Pro Arg Ser Gly Glu Gln Ala Arg Gly Gln Gly Gln Pro 340 | 345 | 350 |
| Pro Arg Ser Pro Ala Gly Gly Trp Glu Leu 355 | 360 | |
| <210> 62 | | |
| <211> 391 | | |
| <212> PRT | | |
| <213> Homo sapiens | | |
| <400> 62 | | |
| Met Arg Gln Leu Cys Arg Gly Arg Val Leu Gly Ile Ser Val Ala Ile 5 | | |
| Ala His Gly Val Phe Ser Gly Ser Leu Asn Ile Leu Leu Lys Phe Leu 20 | | |
| Ile Ser Arg Tyr Gln Phe Ser Phe Leu Thr Leu Val Gln Cys Leu Thr 35 | | |
| Ser Ser Thr Ala Ala Leu Ser Leu Glu Leu Leu Arg Arg Leu Gly Leu 50 | | |
| Ile Ala Val Pro Pro Phe Gly Leu Ser Leu Ala Arg Ser Phe Ala Gly | | |

| 65 | 70 | 75 | 80 |
|---|-----|-----|-----|
| Val Ala Val Leu Ser Thr Leu Gln Ser Ser Leu Thr Leu Trp Ser Leu | | | |
| 85 | 90 | 95 | |
| Arg Gly Leu Ser Leu Pro Met Tyr Val Val Phe Lys Arg Cys Leu Pro | | | |
| 100 | 105 | 110 | |
| Leu Val Thr Met Leu Ile Gly Val Leu Val Leu Lys Asn Gly Ala Pro | | | |
| 115 | 120 | 125 | |
| Ser Pro Gly Val Leu Ala Ala Val Leu Ile Thr Thr Cys Gly Ala Ala | | | |
| 130 | 135 | 140 | |
| Leu Ala Gly Ala Gly Asp Leu Thr Gly Asp Pro Ile Gly Tyr Val Thr | | | |
| 145 | 150 | 155 | 160 |
| Gly Val Leu Ala Val Leu Val His Ala Ala Tyr Leu Val Leu Ile Gln | | | |
| 165 | 170 | 175 | |
| Lys Ala Ser Ala Asp Thr Glu His Gly Pro Leu Thr Ala Gln Tyr Val | | | |
| 180 | 185 | 190 | |
| Ile Ala Val Ser Ala Thr Pro Leu Leu Val Ile Cys Ser Phe Ala Ser | | | |
| 195 | 200 | 205 | |
| Thr Asp Ser Ile His Ala Trp Thr Phe Pro Gly Trp Lys Asp Pro Ala | | | |
| 210 | 215 | 220 | |
| Met Val Cys Ile Phe Val Ala Cys Ile Leu Ile Gly Cys Ala Met Asn | | | |
| 225 | 230 | 235 | 240 |
| Phe Thr Thr Leu His Cys Thr Tyr Ile Asn Ser Ala Val Thr Thr Ser | | | |
| 245 | 250 | 255 | |
| Leu Phe Ile Ala Gly Val Val Val Asn Thr Leu Gly Ser Ile Ile Tyr | | | |
| 260 | 265 | 270 | |
| Cys Val Ala Lys Phe Met Glu Thr Arg Lys Gln Ser Asn Tyr Glu Asp | | | |
| 275 | 280 | 285 | |
| Leu Glu Ala Gln Pro Arg Gly Glu Glu Ala Gln Leu Ser Gly Asp Gln | | | |
| 290 | 295 | 300 | |
| Leu Pro Phe Val Met Glu Glu Leu Pro Gly Glu Gly Asn Gly Arg | | | |
| 305 | 310 | 315 | 320 |
| Ser Glu Gly Glu Ala Ala Gly Gly Pro Ala Gln Glu Ser Arg Gln | | | |
| 325 | 330 | 335 | |
| Glu Val Arg Gly Ser Pro Arg Gly Val Pro Leu Val Ala Gly Ser Ser | | | |
| 340 | 345 | 350 | |
| Glu Glu Gly Ser Arg Arg Ser Leu Lys Asp Ala Tyr Leu Glu Val Trp | | | |
| 355 | 360 | 365 | |
| Arg Leu Val Arg Gly Thr Arg Tyr Met Lys Lys Asp Tyr Leu Ile Glu | | | |
| 370 | 375 | 380 | |

Asn Glu Glu Leu Pro Ser Pro
385 390

<210> 63
<211> 442
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 220,391,428
<223> n = A,T,C or G

<400> 63
atagaacga ctgatgtt tattcgatga aatagggtg ggggtgtgc agccctagtc 60
ccacattcga tgggtgttg actgagttaa cagcaaaatg ggatgcaaaa gttccctgat 120
tggagccccc cggatccgg ttctggattt gctggccact tactctatga ctggggcat 180
gtcaactgtca tggctcaatgtt ttccttgcat gcacagtgtt ttatggata gttccagetc 240
tgacatgtca ggattatgtt atactgtcaa tcaagactag gtttggctta agcacatgg 300
ctgaaacctt actcgggtca tgacatattt ttctccggat ggggagatgg cagctgtca 360
gtggcaaggc tggcccttcaa agctgtccat nccacgcccc ggtgtctgtt ggttcctt 420
ccctcgtncc cgaatttgc 442

<210> 64
<211> 456
<212> DNA
<213> Homo sapiens

<400> 64
cttcaaccat aaaacaaag ggctctgatt gcttagggg ataagtgatt taatatccac 60
aaacgtcccc actccaaaat gttaacttat tctgtgattt aacttttctt ctaatgtga 120
atctttctgt ttttttctct taaggagggaa agttaaagga cactacagggt cataaaaaac 180
aagttggcca aggactcatt aactgtctta tatttttact gccactaaac tgccgttatt 240
tctgtatgtc ttctatccca aacagacgtt cactgccact tggtaaatgtt aaggatgtaaa 300
cgaggatata taactgttcc agtggacaga ttttgtaag tgccctgtt ttttagactt 360
taaagtttatc acattttgtt gacttctgac attccactt cctagttat agggaaagatc 420
tggttatgtt gttttttttt aaaaatgtgcc aatgcc 456

<210> 65
<211> 654
<212> DNA
<213> Homo sapiens

<400> 65
aataaatccc agcccttctt ttcttgctgc ttccctagat atttccctcc tttcttctcc 60
agtatttcaact ctctttctgt gagtttgcat ggcctgttta tggttttgcg gtgtttttt 120
ttctgtatgtt ttttttcttc atatccatc atatgtctatataa ggtttccat attttagggc 180
aggctttggat atttttctgatc caggccataaa agaaatcgat tttttctgtt aagctgttct 240
tttagtgcgtt ataggctgtt atccaaactt cggaaatgtt tttagacaaa atcccttgc 300
aataaaggcc ggagagagaa attttcataat gctgtttggat ataaaacccat ttccttcaat 360
ggagttgtact gtatgttcat cccctaggatt tcttttaaat aactgagaca cccaaactcg 420
gagatcatatc gaataatgtca aagacacagg tggaaaagat ccagatatta tcttcagttac 480
cgatgttttc atactgtccatc cttccaaacgc atgtttaaatgtt acatcagatc cttttggatc 540
acaccccaaga gcccaaaatc ctttttatttttataat tttgtaaatgtt gttttccacag atttccatgt 600

acaagtaqat tggggccccc atcaagttcg gccccctctc cagtctttt qaac 654

<210> 66

<211> 592

<212> DNA

<213> Homo sapiens

<400> 66

```

ttttttttt tttttttatt gggataaat ttatcaaaaa acatgtcata caattccac 60
aatagagaca ttttaaaatc agaatcac tctgttcata atataaaaatcc cccagtggaa 120
agtccctta aaccactta ttttttttttgc tttccatggata attttaaaatc tttttccatgg 180
aattccctta aaccttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 240
tttttttttta agacatgtc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 300
gtttccctttt gtcattttttca ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 360
tttttttttca caacgggtca gggccggcgc ctccaaagg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 420
gaaaacttcc caacatataatc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 480
gtttatccatc agcgattttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 540
gccaatccca aaccggatcc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 600

```

<210> 67

<211> 469

<212> DNA

<213> Homo sapiens

<220>

<221> misc feature

<222> 245_298_314_339_424_440_465

6222> 245,258,514,55
6233> P = A T C or S

54002 67

<210> 68

<211> 510

<212> DNA

(212) DNA

52203

<221> misc feature

<221> misc_1e
<222> 134 163

~~6222~~ 424, 402

1400 68


```

<210> 73
<211> 485
<212> DNA
<213> Homo sapiens

<400> 73
tttttttat ttttagata ttttattttt atgcaaata aattttctata tatgtgaac 60
tgttaaagg gagattagg aactcttcc ttcttcctcg tttttttttc tggtttttct 120
ttttttattt atttttggat tatagatgt ctcttcgtt gcaatgtca atgtttccaca 180
tcttcgttcc acggccctggc ttgttccag gggttttagt gagtgcttc tgtaaggca 240
ttaataatac acggccctggc ttgttccag actccaaatac agggtgtcat acatcaaggaa 300
gcaaggccct gacttttagct ccggaaacgc ctccctttgtt gtcttgatata ttgcactg 360
acatgaccac tgccgtcaca gccagggtt ggacagctga acagctttt tatggctgtt 420
ttcacggaa ctgcgaaccc ttggacccg gtgcgtatcc getttttccca ggtgtcaac 480
ttccat                                         485

<210> 74
<211> 338
<212> DNA
<213> Homo sapiens

<400> 74
ttttttgatt atttcgaga ttatttgc aa gttaatgtc tttttttttt gatattccct 60
aacatgaagg taataaaactt taacgttcca cttaaaaaaaa caaaaaacca acaacgaaa 120
ataagaattt saccagaaag ctatagtttgg tttttttttt cttttttttt cttttttttt 180
taatggataca ttcttgcata ttggatattttt tcaaaaatttta aaggcttttta gtctttaaaag 240
tataaaagaaac ttgtatgttcc ttcttagtcaa cttaatctttt cttttttttt cttttttttt 300
gacagtttca atacttttcc ttggatattttt atatagtcc                                         338

<210> 75
<211> 334
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 265
<223> n = A,T,C or G

<400> 75
agccggccgg cggggccggc aacagtttcta cctgtcttgc ggaaacctgc tcagccccga 60
caatgttgc cggaaacagg cagagggaaac ctatgaaaat atccaggccc agtcaatgg 120
cacatccctc ttacaagccca tccggaaatcc aacagtttgc ttggatattttt cttttttttt 180
ccggccgttcc ttggatattttt cttttttttt cttttttttt cttttttttt cttttttttt 240
cctttttttt cttttttttt cttttttttt cttttttttt cttttttttt cttttttttt 300
acaatcttagc atggggaaaa aagggtttgtt atat                                         334

<210> 76
<211> 248
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> 32,33
<223> n = A,T,C or G

```

<400> 76
 gataggcata aacgtgttta ttaagtaaaa cnnatccctt aaaaataaaa aagggaaagcc 60
 tgatatataa tgaagttgt gattcaacta gccagaattt attctgactt gcaccaaacc 120
 acacaaaatc tttaaaaagt ctatgttc gtatgtctaa tggacactcc agatgtgtt 180
 ctggaaatcc attgcaagag ctccaaatcc ctactttcg aaggatggg gatcaagatg 240
 agggttgt 248

<210> 77
 <211> 515
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> 395,476
 <223> n = A,T,C or G

<400> 77
 atgtagaaac agcatcaagc tggttctctc taccgtctt gatagaataaaa aaaaataaaa 60
 taaaaatgg aattgcagaa aagctaagag gtttttagtt tttgttttt gttttcttc 120
 caccaggca ttatggaaaa ggatgttagt agtctgtttt attttagttt caatctgggt 180
 ttgtacacaa gcaaaaaaagc aatgttgat tttcaggttag accttcatgc agacatgca 240
 aaccaactgt ctgcgtgtg aggagocatg gggagcttc cgaagggtt tccaggcoagt 300
 ggctaatgg aatggatgtac tactcgttag ccctgtcgac cgtatgttaa ggtgtgccaa 360
 ggatatctat cagccccatc gagaatatga aacaanagtgc tgagattcta ttacctaag 420
 taacaagaa accgtaagca acacgactga cagcccaag ggaacactgg aatggngggg 480
 tgaatgtgt ctcgtttagc acccccccaat ctgcg 515

<210> 78
 <211> 532
 <212> DNA
 <213> Homo sapiens

<400> 78
 cctgttgt tatagtttat tactgtcata gctaagaaaa ggcagtcgtat ttcaacataa 60
 tccatatactca tggtaaaat ctc当地actt aggatatactca tggtaaaat tggtaatttat 120
 aacctggta gtattctaaa caaaatattt acaatccatt agtgcaccta aaatctttat 180
 aagctgtatc actgtttaa caaaatccaca cgtatgttago aaaatgtatc acagatgtat 240
 ttatataatc ttatataacce ggcatggact aaaaatatac gataaaaattt gggcaatata 300
 aaaggagggt tgcattttttt atatttttt cattgtatcattt cattgtatcattt acaaaaaggc 360
 taataaaaaaa atataatgtt aaggcaatgtt caaaatccaa agataacatgt tgcccacgg 420
 gggagggtt gggagggtt gtcacaatca agggggggca caaaacagcc ttcatgtttaa 480
 ttgtttttat taagggggaa gtcattgtt gatagtctttt acatctttt at 532

<210> 79
 <211> 431
 <212> DNA
 <213> Homo sapiens

<400> 79
 gggtataaaca aatgtgttcc aacctttttt ctgataatag ccagtaaattt tgcaaaagaga 60
 ggagacaac tggatgttga tacataaaaaa cacctgttcc cactttttttt ttttaatatac 120
 tatataatgtt actgttattt atttttttttt atgaagacag caaaatattt cacataaaaa 180
 tatcttacag aaatcattat tttttttttt aagaaaaaccatttataactaa gttiacagg 240
 aaaaatataa acggaaaaatc ctccctggaa cacttttttga actgaggattt tcactttccata 300
 gttttttttt gtaaacatgtt ctcagggttcc ttttttttttt gtaggttccatc ttatcaatc 360
 gaattacatgtt ctcagggttcc ttatcaatc ttatcaatc ttatcaatc 420
 tcagggttcc ttatcaatc 431

tgaacacagt caaataacct gagaagtgc agatggaaaa gcaacagaat gcaaggcacct 120
 tctaaggct gtaatctt gattactgt gaaaagttc agaacatcat agactcttac 180
 tgccacatgt tcatacacc ctggaaaata acagttaaat tcataatgtat acacatataat 240
 atgaatacac actcatgcat gcacactgtc ttccacacacc cttccctcacc acttaacccg 300
 agttacatc atgttctca gatatgtcat tgcatgttgg tttttctgc atctcaacta 360
 agtgcacggc ctggcgcgt tgacattaat tatgaaat gtaaaacaacc aagcaggcac 420
 attttgggg tgagttttaa gaaatctgtg acctgaaaga aattctgtgg ggactgtctg 480
 ggttattcccg ttatccgt gattatattc tgtttttagg tcttgaccta ttttaagct 540

<210> 84
 <211> 559
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> 493,499,506,517,537,550,559
 <223> n = A,T,C or G

<400> 84
 gtgtttctgt ctgtttttac tcggacaatg cttatttac agcggattt acaaataag 60
 cttatattt cacaatccaa gaaacaccat cacaggagg ttgttaggtg gctgtgtgt 120
 ttccaaaaca gcaaaatagg ttcttccat ccaacccctt ttctcttgtt agatgtgg 180
 gtgtttctgt gggcttcgtc ttctctgcg cacagaaact ggacggactg gtcctctg 240
 aegggccctt gtcgaaggaa atggcgcacg attttgcaca caggtaaca ggctccctca 300
 taacaacact gtgcattttt gtgtcatttt gtttattgtt cactgagggtt ttgcacccctc 360
 agtctttgtt gggaaaatc ggggtgcacg attttgcgtg cacaagaaga tggatgtcc 420
 atgttcgttgg agggacacacg ggcagccccca gccagatccc actggccat gcaggcata 480
 gcagtagaaaa ctnaacgtnc cactntgtaa caggotncaa gacaccaatt ccggcancat 540
 gggaaagaaan taaacccctt 559

<210> 85
 <211> 2466
 <212> DNA
 <213> Homo sapiens

<400> 85
 agttggcccg agtcttggtc gcagagacag gaaatgttaa ttctcagcgt 60
 gtcacccccc acatcttcgc ttactgtctc ggccggccat ctggctctgg ggaccggcg 120
 gccccttcgtc ctccggaaa agctctgtca agtctccggta tattgtatcg gatttattca 180
 cagacagatc ctgcagggtg tgcagtgcc acgtgtatcc cgaatccgcg cgtgtggccc 240
 actacagagatc tggaaaatcat gcaagaaatc tccgactgtt ttatcatgtt caccggcagg 300
 atggaggatc ttctccggaa aggtccggat cagaaaaatgg aatgtatgtcc gacatgtgg 360
 ataacaacaa gtgtgcacaat ctctgcaca gttttaaaact cttgtcttaga gagaagaccc 420
 cccattatca aggcaaaatc cagcggcaaga gttttaaaact cttgtcttaga gagaagaccc 480
 cattaaagac cacagcaaccc cccctggggcc cactttaagcc cccacggatg gacactgtcc 540
 cgggtttccgc atctccctat caaaaagatc attcacacatc atactgtgg ctctgtgcag 600
 cctggtttaa taaccctctg atggcccgac aacattatgt tggcaagaaa cacaataaa 660
 atgcggccatc agttgttttgg ttagaacaac tggggacaaad ctggatatgg ggggaaactg 720
 gaggtctgtgg ggcgtttaatc agatgttccca ttcgttgcgtt ctccctaaac tcaatagaac 780
 agtatcatgc ccatctgtaa ggtatctaaacc accagacccaa ctgttggaaat aatgtgtgg 840
 agcatcaatc aacgacataac aaaaatcatc tagtattttt ctggcgttgg gaattgttta 900
 tcaaccccca gaggaggctt ctttctgtaa caataaaatc ttctttaaaat gattccacaga 960
 ttccatcatc actgtatctg attttttggaa atgaatgttgg tttttttttt ctttttttttt 1020
 tttttaattt tggggtaatgt ttagtattt ggtatgtttt ttaaattttt tctctgtataac 1080
 atattttagca catgttctaa attataatcc tatacggaaac agttggatca ttattttttttt 1140
 tggaaatgttggaaaat tttccaaattt attcctgttgg tttttttttt tttttttttt 1200
 tggtaatcc tcaatgttgcgtt gatgttggaaac cacctctatc cagaatataatc tttttttt 1260

tcatatgtt ggacacagt gagggtaca tgcacagaac tggAACAGAT cactattAGT 1320
gaaaatacc aaatggacaa ataaataccca tgctttccca agccacaggc 1380
ccaggTTTC acatcgcca ataaagacca agggaaatGGG TAAAGGAAGA atttcATCT 1440
tttttctgtt acatTTCAAGG aacgtttct acaaGGTTAGG CAAAGTCCTC CTGCGAATT 1500
gccaaataat acgtttagaa aagaatTTT ctgcgtcat gtatGATCTT ttAGGCAAAA 1560
acgtcttcac acggccTTGAC ctgggtaat tttttcccccaaaaAGCATCC AAAAGAAAGA 1620
ttataaaaaCC cagaacgaga tggaaataaaatTTTTTTTGAGGTTTTGGCT 1680
gaactgtggg cttaatTTGGGG gggatactga tgctttggaa aagaATGTGAGA aatttCTGAA 1740
gaaatggggg ctttggGCTA ggGGGGTCC CCTATTCTT CTGTTCTCA CTGAAAGCTC 1800
actgtcgatc caaAGACTCG TCACTGTGA aagAGACATCG CGATAAAGA aaACAGTCTC 1860
ttttctgtgg gggacGGCTG AGTCGACATG ATGAGCTCT TTCTTAGGT TTAATCTTCTT 1920
tccatgttg cggacatTTG TGCTTGATG CCTGGTAAc AGTGGGGAG TGAGCTTCT 1980
actgtacgtg cttccgttgg accatGGTTG ATGCTGGGG GGTGCGCTG TGTCCCTAAC 2040
gaccatagc aatccatott ctgtgtcag caatttCTGG acacccACTG TTTCCACCA 2100
agactgtgggg tggcaacacG TCACTGTGACA taaaACAAAGG TGACACAGGAGG ATGCAACAGT 2160
tttgtatgg aaggccGCTG TACCTGTGTT CAAATCTGG CACCCATTCC TTGAGCAGGG 2220
cccgccatgg aggacGACGG TGTCCAGGT TGTGCGCTG CAGAGAGACG AAGCCCCACG 2280
aggccacacc tacttacaaG gggaaAGGG GTTGGGTGG GAAGAATCTA TTGTTGTGT 2340
ttggaaAGCAG cacaacGGC CTAAACACTT CTGTGATGG TGTTCCTGG GTATGTTAA 2400
ataaggGTTT atttgtaat tcccgctgtaa aataAGCATT GTCCGAGTAA AAACAGCAGC 2460
aacacac

<210> 86
<211> 408
<212> DNA
<213> *Homo sapiens*

<210> 87
<211> 431
<212> DNA
<213> *Homo sapiens*

<220>
<221> misc_feature
<222> 361,431
<223> n = A,T,C or G

```
<400> 87  
acaacccctt cgggggttc ctgatggct gctctcgaa aaggccatcc taaaataaagg 60  
gggggggtta tagggcagac tgaggagag gacaggcttc gagttactgc tacaggttca 120  
ggtcaactggg ctcccgacga gatcggttt tctccgggtt ctcgagagct ggcgttgtt 180  
ctcatcaaaa ccgtacagcc agctaattgc atggacaaat ttatccatcca cacaatggat 240  
gcacgcgttc cgaggccgtcg ggccgggtt atggatgttg gtggcccaaga ccaacagct 300  
gtgcgagact ctgggcattt cggttttagt atacaaggat tgcttaatgg tcacagtcc 360  
nagaatgtta ttccatgtat tgtagtactt ggtacacac aaagaaaaaa aaaaaaaaaact 420  
tcaatcccccc n 431
```

<210> 88
<211> 385

<212> DNA
<213> Homo sapiens

<400> 88

<210> 89

<211> 272

<212> DNA

<213> Homo sapiens

<400> 89

```
tctttaaaat acatacgaat gttaagagaa aatggccaa acctaataac tacgattgtt 60
gaaaacaata tttaaaaggc aatccatataa atactgtcat aaaaatagtgt ttatcttgtt 120
taactaatgt tacatctttt ttcccaatcc catgttgc aagaatgtttt atgcggcgca 180
tggaaaggcac caaggttgcat gtgttattttt gctttaaaaat atacaaaagaa ttgccttactt 240
tggaaaaaaa atatgttcataa ttttttatgtat at 272
```

<210> 90

<211> 504

<212> DNA

<213> Homo sapiens

<400> 90

<210> 91

<211> 467

<212> DNA

<213> Homo sapiens

<400> 91

```

ttttttttt ttttttttgc ttcttcaca aatagtttac tcgggaaac ctaacagaac 60
taatattttt ttctgtccgt aaaaaaaaat agatcatgt tgatgtgtc acttgcgcg 120
aactccccca gtctttcccgc attcgttcgt ctccctccctc caactctgtt ttatccagc 180
ggggggggaa agatctttt ggttgcggc aatcaacga cttagaaaaaa gagggccctad 240
ctggccaaagg aacaggccctt cccttccttc gccttcgttc cttcgtccgt ttccgttccct 300
ttccctttt ttccttgggg ttcccttcgc cggttaacta tggggacaga cacagctatt 360
caacatgtc ttcggcggc acactccggat gttaaggccacg aaggccggaa gacagggttcc 420
cgatcccaa atctggggaa agatggatgtt aagctttttt ctggat 467

```

<210> 92

<211> 229

<212> PRT

<213> Homo sapiens

<400> 92

| | | |
|---|----|----|
| Met Lys Ser Ser Asp Ile Asp Gln Asp Leu Phe Thr Asp Ser Tyr Cys | | |
| 5 | 10 | 15 |

| | | |
|---|----|----|
| Lys Val Cys Ser Ala Gln Leu Ile Ser Glu Ser Gln Arg Val Ala His | | |
| 20 | 25 | 30 |

| | | |
|---|----|----|
| Tyr Glu Ser Arg Lys His Ala Ser Lys Val Arg Leu Tyr Tyr Met Leu | | |
| 35 | 40 | 45 |

| | | |
|---|----|----|
| His Pro Arg Asp Gly Gly Cys Pro Ala Lys Arg Leu Arg Ser Glu Asn | | |
| 50 | 55 | 60 |

| | | | |
|---|----|----|----|
| Gly Ser Asp Ala Asp Met Val Asp Lys Asn Lys Cys Cys Thr Leu Cys | | | |
| 65 | 70 | 75 | 80 |

| | | |
|---|----|----|
| Asn Met Ser Phe Thr Ser Ala Val Val Ala Asp Ser His Tyr Gln Gly | | |
| 85 | 90 | 95 |

| | | |
|---|-----|-----|
| Lys Ile His Ala Lys Arg Leu Lys Leu Leu Leu Gly Glu Lys Thr Pro | | |
| 100 | 105 | 110 |

| | | |
|---|-----|-----|
| Leu Lys Thr Thr Ala Thr Pro Leu Ser Pro Leu Lys Pro Pro Arg Met | | |
| 115 | 120 | 125 |

| | | |
|---|-----|-----|
| Asp Thr Ala Pro Val Val Ala Ser Pro Tyr Gln Arg Arg Asp Ser Asp | | |
| 130 | 135 | 140 |

| | | | |
|---|-----|-----|-----|
| Arg Tyr Cys Gly Leu Cys Ala Ala Trp Phe Asn Asn Pro Leu Met Ala | | | |
| 145 | 150 | 155 | 160 |

| | | |
|---|-----|-----|
| Gln Gln His Tyr Asp Gly Lys Lys His Lys Lys Asn Ala Ala Arg Val | | |
| 165 | 170 | 175 |

| | | |
|---|-----|-----|
| Ala Leu Leu Glu Gln Leu Gly Thr Thr Leu Asp Met Gly Glu Leu Arg | | |
| 180 | 185 | 190 |

| | | |
|---|-----|-----|
| Gly Leu Arg Arg Asn Tyr Arg Cys Thr Ile Cys Ser Val Ser Leu Asn | | |
| 195 | 200 | 205 |

| | | |
|---|-----|-----|
| Ser Ile Glu Gln Tyr His Ala His Leu Lys Gly Ser Lys His Gln Thr | | |
| 210 | 215 | 220 |

Asn Leu Lys Asn Lys
225

<210> 93

<211> 2327

<212> DNA

<213> Homo sapiens

<400> 93

gggagcgaaa accaacgtgt tcggtgacag acccccagcgc cgactgagcc tctaaagcga 60

<210> 94

<211> 2370

212 DND

<212> DNA

52202

<221> misc feature

<222> 741,1195,1683,2360

<223> P = A-T-C or G

<400> 94

ggggcgggag ctgaagtgt a gaggactc tt tagacagca gaaaggaaa gccgtt gaga 540
 agttcccttc aaaccttccc tgccctc ctaatccaa cc tccactccc ttctccaaaaa 600
 gttaaaaagg a aagccaaagg t tgccacgctc ccctgttct actcaataaa tacttcttct 660
 aactccgcac cggaaaaaca gaaaaaaaaa actaattcc ttcccaatata taggacttag 720
 aaaaggctca ggtcccgcaaa yttgaatttt agccatgggg aatcaaaaata gtaggagcat 780
 tactcttgtt tccttttca aatacccaaa cctcatcctt ctgcgcacgc catgttacc 840
 aacatttgta gtttcaagga cagggtcg tccatctgt ttgcacaatt ctgtaaacaa 900
 gtgtcagct ctaggggaaat gaaggctt ttgtgtgtt atactgaaat agacccatgg 960
 tctacagaca tcccttcata caacgcagg t gacttacty gaaatgtca ttccacccaa 1020
 atotgcaaat gttaaaatgg gacatcgca tgtttaaaat gtggaaacat tgtaggttat 1080
 catgtgtt tccatgtag ttctgttctt ctccctgc aacacggaca ctctggat 1140
 ttccacagcc aggacgttta tgatataac agactagact ccacagggtt aacatctca 1200
 ctttggggca ac tttccaga gataagaag agatcagatc agatgtgtt aatatactca 1260
 gcaaggaggat gtatagata aatggaaat tgatataat gatataaaa cttttttca 1320
 tttaaaaata tattaatgg tcaactttaa aatttgttagt tgccagtat cttttttgg 1380
 aaacacatgg tttttttttt tttttccgtt ctcttaattt tgtaggttcat 1440
 ttgattggaa ccacgttggat tttttttttt tttttttttt tttttttttt 1500
 ttctggccatg tttttttttt tttttttttt tttttttttt tttttttttt 1560
 agttccctcc caatctgggg gatgtttttt tttttttttt tttttttttt 1620
 acatcttcc tttttttttt tttttttttt tttttttttt tttttttttt 1680
 acyaggat tttt aataggggat tttttttttt tttttttttt tttttttttt 1740
 catatgttctt atttttttttt tttttttttt tttttttttt tttttttttt 1800
 agatggat tttttttttt tttttttttt tttttttttt tttttttttt 1860
 gataaaatgg tttttttttt tttttttttt tttttttttt tttttttttt 1920
 ttctttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1980
 ttcccttcc tttttttttt tttttttttt tttttttttt tttttttttt 2040
 aatagaaaaaa aaaaactaaag cttttttttt tttttttttt tttttttttt 2100
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2160
 attagaaatgg agggtatgt tttttttttt tttttttttt tttttttttt 2220
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2280
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2340
 aaaaaaaaaa aaaaaaaaaaam aaaaaaaaaa 2370

<210> 95
 <211> 450
 <212> DNA
 <213> Homo sapiens

<400> 95.
 atgttccatca acatttgtat ttcaaggac aggtgtgtt ccatctgtt ttgcaatatt 60
 gttaaaaacaa tgctcagtc tagggaaatg aaggcttctt tgctgtgtt tactgaaata 120
 gacccctttt ctacagacat ccctccatca aacccgttgg atttcaactgg aatgtgtt 180
 ttccacaaaaa tttttttttt tttttttttt tttttttttt tttttttttt 240
 gttaggttac atgtgttgc tttttttttt tttttttttt tttttttttt 300
 ttctgtgtt tttttttttt tttttttttt tttttttttt tttttttttt 360
 aacgttccatca ttggggccaa ctggcaggat tttttttttt tttttttttt 420
 aatatctca cagaggatgt tttttttttt tttttttttt tttttttttt 450

<210> 96
 <211> 149
 <212> PRT
 <213> Homo sapiens

<400> 96
 Met Ser Thr Asn Ile Cys Ser Phe Lys Asp Arg Cys Val Ser Ile Leu
 5 10 15

Cys Cys Lys Phe Cys Lys Gln Val Leu Ser Ser Arg Gly Met Lys Ala
20 25 30

Val Leu Leu Ala Asp Thr Glu Ile Asp Leu Phe Ser Thr Asp Ile Pro
35 40 45

Pro Thr Asn Ala Val Asp Phe Thr Gly Arg Cys Tyr Phe Thr Lys Ile
50 55 60

Cys Lys Cys Lys Leu Lys Asp Ile Ala Cys Leu Lys Cys Gly Asn Ile
65 70 75 80

Val Gly Tyr His Val Ile Val Pro Cys Ser Ser Cys Leu Leu Ser Cys
85 90 95

Asn Asn Gly His Phe Trp Met Phe His Ser Gln Ala Val Tyr Asp Ile
100 105 110

Asn Arg Leu Asp Ser Thr Gly Val Asn Val Leu Leu Trp Gly Asn Leu
115 120 125

Pro Glu Ile Glu Glu Ser Thr Asp Glu Asp Val Leu Asn Ile Ser Ala
130 135 140

Glu Glu Cys Ile Arg
145